

การศึกษาการแยกความหมายของคำหลายความหมายในภาษาไทยโดยใช้วิธีการวิเคราะห์ความหมาย
แบบแฝง

นางสาวนัชชา ถิระสาโรช

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาอักษรศาสตรดุษฎีบัณฑิต
สาขาวิชาภาษาศาสตร์ ภาควิชาภาษาศาสตร์
คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2559
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A STUDY OF WORD SENSE DISCRIMINATION IN THAI USING LATENT SEMANTIC
ANALYSIS

Miss Nutchra Tirasaroj

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Linguistics

Department of Linguistics

Faculty of Arts

Chulalongkorn University

Academic Year 2016

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การศึกษาการแยกความหมายของคำหลายความหมายใน
	ภาษาไทยโดยใช้วิธีการวิเคราะห์ความหมายแบบแฝง
โดย	นางสาวนัชชา ธีระสาโรช
สาขาวิชา	ภาษาศาสตร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.วิโรจน์ อรุณมานะกุล

คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญาโท

.....คณบดีคณะอักษรศาสตร์
(รองศาสตราจารย์ ดร.ก้องกัญญา เทพกาญจนา)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร.ก้องกัญญา เทพกาญจนา)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.วิโรจน์ อรุณมานะกุล)

.....กรรมการ
(รองศาสตราจารย์ ดร.โชติรัตน์ รัตนมัทธนะ)

.....กรรมการ
(อาจารย์ ดร.วรรณชัย คำภีระ)

.....กรรมการภายนอกมหาวิทยาลัย
(ดร.เทพชัย ทรัพย์นิธิ)

นัชชา ถิระสาโรช : การศึกษาการแยกความหมายของคำหลายความหมายในภาษาไทยโดยใช้วิธีการวิเคราะห์ความหมายแอบแฝง (A STUDY OF WORD SENSE DISCRIMINATION IN THAI USING LATENT SEMANTIC ANALYSIS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. วิโรจน์ อรุณมานะกุล, 106 หน้า.

วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์เพื่อพัฒนาระบบการแยกนัยความหมายของคำหลายนัยความหมายโดยใช้แนวทางการวิเคราะห์ความหมายแอบแฝง และศึกษาเปรียบเทียบประสิทธิภาพของระบบในการแยกนัยความหมายของคำหลายนัยความหมายที่อยู่ในประเภทของคำเดียวกันและต่างกัน รวมถึงระบุบริบทที่เหมาะสมในการแยกนัยความหมายของคำที่ใช้ศึกษา คำที่ใช้ศึกษาได้แก่ คำนาม คือ คำว่า เสี่ยง และ หัว และคำกริยา คือ คำว่า บอก และ คิด โดย เสี่ยง และ บอก เป็นตัวแทนของคำที่มีนัยความหมายน้อย และ หัว และ คิด เป็นตัวแทนของคำที่มีนัยความหมายมาก ในการศึกษาได้ใช้คำบริบทตำแหน่งต่าง ๆ ในการช่วยแยกความหมาย ผลการศึกษาพบว่า ระบบที่ใช้คำบริบทที่อยู่ติดกับคำเป้าหมายและมีกรอบหน้าต่างไม่มากจะสามารถแยกความหมายได้ดีกว่าระบบที่ใช้กรอบหน้าต่างมาก และระบบสามารถแยกความหมายของคำที่มีนัยความหมายน้อยได้ดีกว่าคำที่มีนัยความหมายมาก เนื่องจากการกระจายตัวของข้อมูลมีน้อยกว่า ส่วนประเภทของคำที่ต่างกันนั้นไม่มีผลต่อประสิทธิภาพของระบบ และตำแหน่งของบริบทไม่มีผลต่อการแยกความหมายของคำนามและคำกริยา เนื่องจากงานวิจัยนี้ใช้เพียงรูปคำเท่านั้น ดังนั้นสิ่งที่ช่วยระบบในการแยกความหมายเป็นหลักคือคำบริบทที่ปรากฏร่วมกับคำเป้าหมายนั้น ๆ เป็นประจำและมีความถี่สูง ในส่วนของค่าความถูกต้องของระบบในงานวิจัยนี้ยังไม่ได้เท่าใดนัก ได้ความถูกต้องระหว่าง x-y สาเหตุน่าจะมากจากการใช้เพียงรูปคำ และจำนวนข้อมูลที่ใช้มีไม่มาก

ภาควิชา ภาษาศาสตร์

ลายมือชื่อนิสิต

สาขาวิชา ภาษาศาสตร์

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ปีการศึกษา 2559

5480510622 : MAJOR LINGUISTICS

KEYWORDS: WORD SENSE DISCRIMINATION / POLYSEMY / LATENT SEMANTIC ANALYSIS

NUTCHA TIRASAROJ: A STUDY OF WORD SENSE DISCRIMINATION IN THAI USING
LATENT SEMANTIC ANALYSIS. ADVISOR: ASSOC. PROF. WIROTE
AROONMANAKUN, Ph.D., 106 pp.

The main purpose of this study is to develop Thai word sense discrimination system using Latent Semantic Analysis as well as comparing the performance of systems when discriminating senses of polysemous words in the same and different word class and indicating the appropriate context help distinguishing the senses. Words used in this study are nouns, /sieng4/ 'sound' and /hua4/ 'head', and verbs, /bɔ̌k1/ 'tell' and /tɪt1/ 'stick'. /sieng4/ and /bɔ̌k1/ are the representations of few-meaning words whereas /hua4/ and /tɪt1/ are the representations of multi-meaning words. Contexts are the clues that help discriminating the senses in this study. The results show that the systems using the small window size tend to distinguish the word senses better than those using the large window size. Moreover, the systems can discriminate the senses of few-meaning words better than those of multi-meaning words because of a small dispersion of data. The different part of speech does not affect the performance of the systems as well as context position does not have an effect on discriminating the senses of nouns and verbs. Since this study uses only word forms, the key that help discriminating the senses is words that co-occur with the target words high frequently. The results of word sense discrimination in this study is not good. Accuracy is ranged from x to y. This could cause from the use of only word forms and the low number of training data.

Department: Linguistics

Student's Signature

Field of Study: Linguistics

Advisor's Signature

Academic Year: 2016

กิตติกรรมประกาศ

ผู้วิจัยต้องขอขอบพระคุณ รองศาสตราจารย์ ดร. วิโรจน์ อรุณมานะกุล อาจารย์ที่ปรึกษาวิทยานิพนธ์เป็นอย่างสูง ที่ได้ให้คำแนะนำและความช่วยเหลือในการทำวิจัย ตลอดจนปรับแก้วิทยานิพนธ์ฉบับนี้จนสำเร็จลุล่วงไปด้วยดี และผู้วิจัยขอขอบพระคุณ รองศาสตราจารย์ ดร. กิ่งกาญจน์ เทพกาญจนา รองศาสตราจารย์ ดร. โซติรัตน์ รัตนมัทธนะ อาจารย์ ดร. วรรณชัย คำภีระ และ ดร. เทพชัย ทรัพย์นิธิ กรรมการสอบวิทยานิพนธ์ที่ได้ให้ข้อชี้แนะและเสียสละเวลาในการตรวจแก้วิทยานิพนธ์ฉบับนี้ให้มีความสมบูรณ์มากยิ่งขึ้น

ขอขอบพระคุณ คณาจารย์ภาควิชาภาษาศาสตร์ทุกท่านที่คอยให้คำแนะนำและประสิทธิ์ประสาทความรู้ด้านภาษาศาสตร์แก่ผู้วิจัย

ขอขอบพระคุณบิดา มารดา คุณภักติน และคุณธมน ธีระสาโรช ที่ห่วงใย ให้กำลังใจ และการสนับสนุนที่ดีตลอดมา คุณสมบัชร ธีระสาโรช ที่ให้โอกาสผู้วิจัยได้ศึกษาต่อและสนับสนุนค่าใช้จ่ายต่าง ๆ ตลอดระยะเวลาการศึกษา คุณธารทอง แจ่มไพบูลย์ คุณคเชนทร์ ตัญศิริ คุณฉลธิกา มหาพูนทอง ที่ให้ความช่วยเหลือ เป็นกำลังใจ และเป็นพี่ที่ปรึกษาที่ดีมาโดยตลอด ขอขอบคุณ คุณจินดา โคกครุฑ คุณพรนฤมล เรียบร้อยเจริญ คุณรุ่งตะวัน ชนะศึก คุณอิสริย์ เพื่องเจริญ คุณสุดาร์ตน์ ดุลสวัสดิ์ ที่เป็นกำลังใจให้ผู้วิจัย รวมทั้งขอขอบคุณเพื่อน ๆ พี่ ๆ น้อง ๆ และเจ้าหน้าที่ภาควิชาภาษาศาสตร์สำหรับการสนับสนุนและช่วยเหลือในด้านต่าง ๆ เสมอมา

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	5
1.3 สมมติฐานของการวิจัย	5
1.4 ขอบเขตของการวิจัย	6
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	6
1.6 วิธีดำเนินการวิจัย.....	6
1.7 เครื่องมือที่ใช้ในการวิจัย.....	7
บทที่ 2 ทบทวนวรรณกรรม.....	8
2.1 คำหลายความหมาย (polysemy).....	8
2.2 ทฤษฎีหรือแนวคิดที่เกี่ยวข้องกับคำหลายความหมาย	10
2.2.1 ทฤษฎีต้นแบบ (prototype theory)	10
2.2.2 ทฤษฎีคำหลายความหมายอย่างมีหลักการ (principled polysemy)	14
2.3 การศึกษาความหมายในทางภาษาศาสตร์คอมพิวเตอร์.....	16
2.4 แนวทางการศึกษาความกำกวมทางความหมายของคำ.....	17
2.5 Latent Semantic Analysis (LSA)	23
2.6 การศึกษาความกำกวมทางความหมายในภาษาไทย	27
บทที่ 3 วิธีดำเนินการวิจัย.....	29

3.1 วิธีการเลือกคำที่ใช้ในงานวิจัย	29
3.2 วิธีการรวบรวมข้อมูล.....	31
3.3 วิธีการทดลอง.....	34
3.4 วิธีการประเมินประสิทธิภาพของระบบ	38
บทที่ 4 การวิเคราะห์ตามแนวภาษาศาสตร์	39
4.1 ความหมายของ <i>เสียง</i>	39
4.2 ความหมายของ <i>บอก</i>	40
4.3 ความหมายของ <i>หัว</i>	42
4.4 ความหมายของ <i>ติด</i>	49
บทที่ 5 การทดลองแยกความหมายของคำด้วยเครื่อง	58
5.1 กรอบหน้าต่างและบริบท	58
5.2 จำนวนความหมายของคำ	67
5.3 ประเภทของคำ	70
5.4 บริบทที่เครื่องใช้แยกนัยความหมายของคำที่ใช้ในงานวิจัย	72
บทที่ 6 สรุป อภิปราย และข้อเสนอแนะ	75
6.1 สรุปผลการวิจัย	75
6.2 อภิปรายผลการวิจัย	77
6.2.1 คำกริยามีนัยความหมายหลากหลายมากกว่าคำนามหรือไม่	77
6.2.2 ระบบแยกคำที่มีนัยความหมายน้อยได้ดีกว่าคำที่มีนัยความหมายมากหรือไม่	78
6.2.3 ระบบสามารถแยกนัยความหมายของคำนามได้ถูกต้องมากกว่าคำกริยาหรือไม่	79
6.2.4 บริบทที่เหมาะสมในการแยกนัยความหมายของคำนามและคำกริยา	80
6.3 ปัญหาที่พบ	81
6.4 ข้อเสนอแนะ	81

รายการอ้างอิง	83
ภาคผนวก.....	88
ประวัติผู้เขียนวิทยานิพนธ์	106

สารบัญตาราง

ตารางที่ 2.1 รูปแบบเมทริกซ์ของ LSA (Kontostathis and Pottenger, 2002)	24
ตารางที่ 2.2 ค่าใหม่ของค่าจากตารางที่ 1 เมื่อผ่านการแยกองค์ประกอบด้วย SVD (Kontostathis and Pottenger, 2002).....	25
ตารางที่ 3.1 ความหมายของ เสียง และ บอก ตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554	30
ตารางที่ 3.2 ความหมายของ หัว และ ติด ตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554.	30
ตารางที่ 3.3 ความถี่ของคำบริบททางซ้ายและขวา 3 ตำแหน่ง 7 คำแรกของ เสียง	33
ตารางที่ 3.4 ความถี่ของคำบริบททางซ้ายและขวา 3 ตำแหน่ง 7 คำแรกของ บอก.....	33
ตารางที่ 3.5 ความถี่ของคำบริบททางซ้ายและขวา 3 ตำแหน่ง 7 คำแรกของ หัว	33
ตารางที่ 3.6 ความถี่ของคำบริบททางซ้ายและขวา 3 ตำแหน่ง 7 คำแรกของ ติด	34
ตารางที่ 3.7 จำนวนตัวอย่างข้อมูลของ เสียง หัว บอก ติด.....	34
ตารางที่ 3.8 การนับความถี่ของคำปรากฏรวม	36
ตารางที่ 4.1 จำนวนข้อมูล เสียง.....	40
ตารางที่ 4.2 จำนวนข้อมูลของ บอก	42
ตารางที่ 4.3 ความหมายของ หัว จากพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554 และจาก งานวิจัยนี้	42
ตารางที่ 4.4 ความหมายของ หัว ที่วิเคราะห์ได้และจำนวนตัวอย่างข้อมูล.....	47
ตารางที่ 4.5 ความหมายของ หัว และจำนวนตัวอย่างข้อมูลที่ใช้ในงานวิจัยนี้.....	48
ตารางที่ 4.6 ความหมายของ ติด จากพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554 และจาก งานวิจัยนี้	49
ตารางที่ 4.7 ความหมายของ ติด ที่วิเคราะห์ได้และจำนวนตัวอย่างข้อมูล.....	55
ตารางที่ 4.8 ความหมายของ ติด และจำนวนตัวอย่างข้อมูลที่ใช้ในงานวิจัยนี้.....	56
ตารางที่ 5.1 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าต่างที่ต่างกันของคำว่า เสียง.....	58
ตารางที่ 5.2 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าต่างที่ต่างกันของคำว่า บอก	59

ตารางที่ 5.3 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าที่ต่างกันของคำว่า หัว	..60
ตารางที่ 5.4 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าที่ต่างกันของคำว่า ดิด	..60
ตารางที่ 5.5 รายละเอียดการจัดกลุ่มความหมายของ เสียง ด้วยคำบริบททางซ้าย 5 ตำแหน่ง	..62
ตารางที่ 5.6 รายละเอียดการจัดกลุ่มความหมายของ บอก ด้วยคำบริบททางซ้าย 2 ตำแหน่ง	..62
ตารางที่ 5.7 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางซ้าย 2 ตำแหน่ง	62
ตารางที่ 5.8 รายละเอียดการจัดกลุ่มความหมายของ ดิด ด้วยคำบริบททางขวา 3 ตำแหน่ง	63
ตารางที่ 5.9 ความถี่ของคำบริบททางซ้าย 2 ตำแหน่ง 5 คำแรกของ บอก	64
ตารางที่ 5.10 ความถี่ของคำบริบททางซ้าย 2 ตำแหน่ง 10 คำแรกของ หัว	65
ตารางที่ 5.11 รายละเอียดการจัดกลุ่มความหมายของ บอก ด้วยคำบริบททางซ้าย 2 ตำแหน่ง	66
ตารางที่ 5.12 รายละเอียดการจัดกลุ่มความหมายของ บอก ด้วยคำบริบททางซ้าย 3 ตำแหน่ง	66
ตารางที่ 5.13 รายละเอียดการจัดกลุ่มความหมายของ บอก ด้วยคำบริบททางซ้าย 5 ตำแหน่ง	66
ตารางที่ 5.14 ค่าความถูกต้องของระบบเมื่อแยกนัยความหมายของคำที่มีนัยความหมายน้อยกับคำที่มีนัยความหมายมาก	67
ตารางที่ 5.15 เปรียบเทียบการกระจายของจำนวนข้อมูลในแต่ละความหมายของ เสียง และ บอก	68
ตารางที่ 5.16 เปรียบเทียบการกระจายของจำนวนข้อมูลในแต่ละความหมายของ หัว และ ดิด	68
ตารางที่ 5.17 รายละเอียดการจัดกลุ่มความหมายของ เสียง ด้วยคำบริบททางซ้าย 2 ตำแหน่ง	69
ตารางที่ 5.18 ค่าความถูกต้องของระบบเมื่อแยกนัยความหมายของคำระหว่างคำนามและคำกริยา	70
ตารางที่ 5.19 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางขวา 3 ตำแหน่ง	...71
ตารางที่ 5.20 รายละเอียดการจัดกลุ่มความหมายของ ดิด ด้วยคำบริบททางขวา 3 ตำแหน่ง	...71
ตารางที่ 6.1 รายละเอียดการจัดกลุ่มความหมายของ ดิด ด้วยคำบริบททางขวา 2 ตำแหน่ง	76

สารบัญภาพ

ภาพที่ 2.1 แผนภาพแสดงการขยายทางความหมายของ radial categories (Evans and Green, 2006).....	12
ภาพที่ 2.2 แผนภาพแสดงความแตกต่างระหว่างความหมายของคำหลายความหมาย.....	13
ภาพที่ 2.3 The bird flew over the yard. ปรับจากงานของ Lakoff (Evans and Green, 2006).....	15
ภาพที่ 2.4 ตัวอย่างของเวกเตอร์คำของคำว่า <i>restaurant</i> และ <i>money</i> (Navigli, 2009)	21
ภาพที่ 2.5 ตัวอย่าง centroid ของคำว่า <i>stock</i> ที่ได้จากเวกเตอร์คำที่ปรากฏอยู่ในบริบทเดียวกัน (Navigli, 2009)	22
ภาพที่ 2.6 ส่วนหนึ่งของกราฟคำปรากฏรวมของคำว่า <i>bar</i> (Navigli, 2009)	23
ภาพที่ 2.7 ลักษณะการแยกองค์ประกอบของเมทริกซ์ด้วย SVD (Clarke, 2007)	25
ภาพที่ 3.1 ขั้นตอนของระบบการแยกความหมายของคำ	35
ภาพที่ 5.1 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าต่างที่แตกต่างกันของ เสียง ...	59
ภาพที่ 5.2 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าต่างที่แตกต่างกันของ บอก....	59
ภาพที่ 5.3 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าต่างที่แตกต่างกันของ หัว	60
ภาพที่ 5.4 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าต่างที่แตกต่างกันของ ดิด	61

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ในงานด้านการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) การศึกษาเรื่องความหมายของคำเป็นหัวข้อหนึ่งที่น่าสนใจและมีการศึกษากันมาอย่างต่อเนื่องในกลุ่มนักภาษาศาสตร์คอมพิวเตอร์ นักวิศวกรรมคอมพิวเตอร์ และนักวิทยาศาสตร์คอมพิวเตอร์ เพราะมีส่วนสำคัญและเป็นพื้นฐานของงานประมวลผลภาษาธรรมชาติอื่น ๆ เช่น การแปลภาษาด้วยเครื่อง (Machine Translation) ปัญญาประดิษฐ์ (Artificial Intelligence: AI) เป็นต้น ทั้งนี้เพราะคำถือเป็นหน่วยพื้นฐานหนึ่งในภาษา ดังนั้นการจะสอนให้คอมพิวเตอร์เรียนรู้ภาษาได้นั้นสิ่งหนึ่งที่จำเป็นต้องสอนให้คอมพิวเตอร์เข้าใจเป็นอันดับแรก ๆ คือความหมายของคำ แต่ทั้งนี้เมื่อพิจารณาดูคำในภาษาแล้วจะเห็นว่าคำเป็นจำนวนมากเป็นคำหลายความหมาย (polysemy) ซึ่งสำหรับมนุษย์แล้ว แม้คำจำนวนมากจะมีหลายความหมาย แต่ก็ไม่ได้เป็นอุปสรรคในการสื่อสารระหว่างกันมากนัก ผู้ส่งสารและผู้รับสารยังคงสามารถเข้าใจความหมายได้ตรงกัน แต่คำหลายความหมายนี้จะปัญหา กับงานด้านการแปล หรือการทำพจนานุกรม (Ravin and Leacock, 2006) และสำหรับคอมพิวเตอร์ ก็เป็นปัญหาเช่นกันว่าจะมีวิธีการใดที่จะสอนให้คอมพิวเตอร์รู้จักความหมายของคำ รวมถึงรู้ว่าควรเลือกใช้ความหมายใดจากความหมายทั้งหมดของคำหลายความหมายเมื่ออยู่ในบริบทต่าง ๆ ซึ่งในงานด้านประมวลผลภาษาธรรมชาติที่ศึกษาเกี่ยวกับความหมายของคำ คือ การศึกษาเรื่องการแก้ปัญหาคำกำกวมทางความหมายของคำ (Word Sense Disambiguation)

การแก้ปัญหาคำกำกวมทางความหมายของคำ คือ การสอนให้คอมพิวเตอร์เรียนรู้ความหมายของคำและสามารถแยกความหมายต่าง ๆ ของคำได้เมื่ออยู่ในบริบทต่าง ๆ สามารถแบ่งเป็นขั้นตอนย่อย 2 ขั้นตอน คือ ขั้นตอนการวิเคราะห์ว่าคำที่ต้องการหาความหมายมีความหมายประจำคำอยู่ที่ความหมาย และขั้นตอนการระบุความหมายให้กับคำนั้น ๆ (Schütze, 1998) ในการแก้ปัญหาคำกำกวมทางความหมายของคำ หากทำในภาษาที่มีการศึกษากันอย่างกว้างขวาง เช่น ภาษาอังกฤษ ก็จะมีทรัพยากรต่าง ๆ มากมายที่สามารถนำมาใช้ได้ เช่น รายการความหมายของคำ คลังข้อมูลที่มีการกำกับความหมายให้กับคำหรือกำกับข้อมูลอื่น ๆ ให้กับคำ เช่น ข้อมูลโครงข่ายคำ

(WordNet) เป็นต้น ซึ่งทรัพยากรเหล่านี้จะมีประโยชน์ในส่วนของ การช่วยวิเคราะห์ว่าคำควรจะมี ความหมายได้ก็ความหมาย แต่สำหรับภาษาที่ขาดแคลนทรัพยากร (under-resourced language) เพียงขั้นตอนแรกที่จะหาว่าคำมีความหมายก็ถือเป็นเรื่องยากแล้ว (Nasiruddin, 2013) ดังนั้นจึงมี การนำเทคนิควิธีอื่นมาใช้ช่วยในการหาความหมายแทน คือ การหาความหมายของคำแบบอัตโนมัติ จากบริบทที่ปรากฏในข้อมูล (Word Sense Discrimination หรือ Word Sense Induction)

สำหรับภาษาไทยก็จัดว่าเป็นภาษาที่อยู่ในกลุ่มขาดแคลนทรัพยากรเช่นกัน เพราะปัจจุบัน คลังข้อมูลที่เปิดให้บุคคลทั่วไปใช้ได้ เช่น คลังข้อมูลภาษาไทยแห่งชาติ (Thai national corpus: TNC) ยังไม่มีข้อมูลทางภาษาใด ๆ เช่น หน้าที่ของคำ (part of speech: POS) ความหมายของคำ เป็นต้น กำกับไว้ แม้ว่าการศึกษาเรื่องความหมายในภาษาไทยจะมีผู้ศึกษากันมาก่อนหน้านี้แล้ว และ อาจจะมีคลังข้อมูลที่ได้กำกับข้อมูลทางภาษาไว้ แต่นั่นก็เป็นคลังข้อมูลที่มีไว้สำหรับศึกษากันภายใน องค์กรเท่านั้น และปัจจุบันรายการความหมายของคำที่สามารถนำมาใช้ได้จริงก็ยังไม่มี แม้จะกล่าวว่า ให้ดึงความหมายที่ปรากฏอยู่ในพจนานุกรมมาสร้างเป็นรายการความหมาย แต่ความหมายของคำที่มี ในพจนานุกรมก็ยังมีปัญหาว่ามีหลายคำที่ความหมายของคำอาจยังไม่ครอบคลุมความหมายที่เราใช้ จริง ดังจะเห็นได้จากคำว่า *หั่ว* และ *เก็บ* ที่ Kanokrattananukul (2001) ได้ศึกษาไว้ ซึ่งพบว่าเมื่อ วิเคราะห์ความหมายของคำจากคลังข้อมูลแล้วได้ความหมายอื่นอีกหลายความหมายนอกเหนือไปจาก ความหมายที่พบในพจนานุกรม ทั้งนี้เนื่องจากความหมายของคำเมื่อเวลาเปลี่ยนไปก็สามารถขยาย ความหมายออกไปได้ โดยเฉพาะคำกริยา ความหมายของคำกริยามักจะขึ้นอยู่กับคำที่ปรากฏร่วมด้วย กล่าวคือ คำกริยาจะมีความหมายหลัก (common core) แล้วความหมายจะสามารถเปลี่ยนไปตาม อาร์กิวเมนต์ (argument) ในประโยค ดังนั้นความหมายของคำกริยาจึงมีลักษณะเป็นปลายเปิด (open-ended meanings) ที่สามารถเพิ่มขึ้นได้ง่าย (Falkum and Vicente, 2015) เช่น คำกริยา *ติด* ใน *เขาติดแอร์* กับ *เขาติดไฟ* ในตัวอย่างแรก *ติด* จะหมายถึง “ติดตั้ง” ส่วน *ติด* ที่สองจะหมายถึง “จุดไฟ” โดย *ติด* จะเปลี่ยนความหมายไปตามกรรมที่ตามมา คือ *แอร์* กับ *ไฟ* เป็นต้น ในขณะที่ ถ้า เป็นคำนาม ความหมายของคำจะสมบูรณ์อยู่ในสิ่งนั้น ๆ อยู่แล้ว เช่น คำว่า *หนังสือ* ก็จะแทนวัตถุที่มี ปกมีเนื้อหาด้านในใช้สำหรับอ่าน ซึ่งความหมายของคำจะขึ้นอยู่กับผู้ใช้ภาษาว่าจะเอาแง่มุมทาง ความหมายใดของวัตถุนั้นมาใช้ (Falkum and Vicente, 2015) เช่น *หนังสือเล่มนี้สวย* กับ *หนังสือ เล่มนี้ดี* จะเห็นว่าประโยคแรก *หนังสือ* จะหมายถึง ส่วนที่เป็นรูปลักษณ์ภายนอก ในขณะที่ถ้าเป็น

หนังสือ ในประโยคหลัง จะหมายถึง เนื้อหาด้านใน เป็นต้น ดังนั้นเมื่อได้วิเคราะห์ความหมายของคำใหม่ จำนวนความหมายของคำจึงมีแนวโน้มเพิ่มขึ้น โดยเฉพาะคำกริยา ที่กระบวนการขยายทางความหมายสามารถทำได้ง่ายกว่าคำนาม

แนวคิดเรื่องการหาความหมายของคำจากบริบทในทางคอมพิวเตอร์นี้ หากมาดูในทางภาษาศาสตร์แล้ว จะพบว่า มีตรงกับแนวคิดของนักภาษาศาสตร์โครงสร้างเช่นกัน Firth (1957) กล่าวว่า “You shall know a word by the company it keeps!” กล่าวคือคำที่เราไม่เคยรู้ความหมายมาก่อน เราสามารถเดาความหมายได้จากคำที่ปรากฏรอบข้าง เช่น ตัวอย่างคำว่า *tezguino* เป็นคำที่สร้างขึ้นมา แต่เมื่อนำไปอยู่ในบริบทต่าง ๆ แล้วก็ทำให้สามารถเดาความหมายของคำได้ เช่น

A bottle of **tezguino** is on the table. Everybody likes **tezguino**.

Tezguino makes you drunk.

We make **tezguino** out of corn.

(Jurafsky and Martin, 2009)

จากตัวอย่างทำให้สามารถเดาได้ว่า *tezguino* น่าจะเป็นเครื่องดื่มชนิดหนึ่ง มีฤทธิ์ทำให้เมา และทำมาจากพืช นอกจาก Firth แล้วนักภาษาศาสตร์โครงสร้างอีกท่านหนึ่งที่มีแนวคิดคล้ายกับ Firth คือ Harris Harris (1968) กล่าวว่า “Words will occur in similar contexts if and only if they have similar meanings.” ตามแนวคิดของ Harris คือคำที่มีความหมายคล้ายกันจะปรากฏอยู่ในบริบทที่คล้ายกันด้วย จะเห็นได้ว่าจากแนวคิดดังกล่าวแสดงให้เห็นว่าบริบทสามารถใช้ช่วยแยกความหมายของคำได้ ดังเช่นตัวอย่างที่ได้ยกไปกรณีคำว่า หนังสือ ใน หนังสือเล่มนี้สวย กับ หนังสือเล่มนี้ดี จะเห็นว่าสิ่งที่ใช้ช่วยแยกความหมายของ หนังสือ คือคำว่า สวย และ ดี ซึ่งเป็นคำบอกคุณลักษณะของหนังสือทำให้รู้ว่าเมื่อใดจะหมายถึงลักษณะภายนอกหรือเนื้อหาของหนังสือ เป็นต้น โดยส่วนใหญ่ในภาษาไทย คำขยายมักจะอยู่ทางด้านขวาดังจะเห็นได้จากตัวอย่างที่ยกมาข้างต้น ดังนั้นบริบทที่น่าจะใช้ช่วยแยกความหมายของคำนามจึงน่าจะเป็นบริบททางขวา ในขณะที่ถ้าเป็นคำกริยา ความหมายของคำกริยาจะขึ้นอยู่กับอาร์กิวเมนต์ในประโยค ดังนั้นบริบทที่จะสามารถช่วยแยกความหมายของคำกริยาได้จึงน่าจะเป็นได้ทั้งบริบททางซ้ายและทางขวา

ด้วยสาเหตุดังกล่าว ผู้วิจัยจึงสนใจศึกษาการหาความหมายของคำแบบอัตโนมัติจากคลังข้อมูล เพื่อพิสูจน์แนวคิดของนักภาษาศาสตร์โครงสร้างดังกล่าวและอีกทั้งยังมียงานวิจัยที่ศึกษา

เรื่องนี้ในภาษาไทยไม่มาก แม้ว่าที่ผ่านมาจะมีงานที่ศึกษาเกี่ยวกับการแก้ปัญหาคำกำกวมในภาษาไทยมาบ้างแล้ว เช่น งานของ Kanokrattananukul (2001) ที่ศึกษาคำว่า หัว และ เกือบ โดยใช้วิธีการตัดสินความหมายโดยดูคำปรากฏรวม ซึ่งความหมายของทั้งสองคำนี้ Kanokrattananukul เป็นผู้วิเคราะห์เอง และงานของ Pongpinigpinyo and Rivepiboon (2005) ใช้วิธี Latent Semantic Indexing (LSI) ในการแยกความกำกวมทางความหมาย และต้องใช้คลังข้อมูลคำว่า หัว และ เกือบ ของ Kanokrattananukul ในการตรวจสอบความถูกต้อง เนื่องจากในภาษาไทยคลังข้อมูลสาธารณะที่มีการกำกับความหมายของคำไว้แล้วยังไม่มี จะเห็นว่าเรายังมีข้อจำกัดในเรื่องทรัพยากรที่จะนำมาใช้ช่วยในการศึกษา แต่ถ้าหากจะต้องให้คนมาวิเคราะห์ความหมายของคำทุกคำในคลังข้อมูลก็ต้องใช้เวลาและแรงคนมาก จึงแทบจะเป็นไปไม่ได้เลยที่จะใช้คนกำกับความหมายของคำในคลังข้อมูล ดังนั้นเทคนิควิธีที่น่าจะเป็นไปได้คือการให้คอมพิวเตอร์ช่วยแยกความหมายของคำให้แทนการใช้แรงงานคน

ในงานวิจัยนี้สนใจวิธี Latent Semantic Analysis ซึ่งเป็นวิธีเดียวกันกับที่ Pongpinigpinyo and Rivepiboon (2005) ใช้ศึกษาเรื่องความกำกวมทางความหมายของคำ เนื่องจาก LSA สามารถหาความหมายของคำและข้อความที่มีความคล้ายคลึงกันได้ถูกต้องค่อนข้างคล้ายกับคำตอบที่มนุษย์ให้มา (Landauer et al., 1998a) และเมื่อเทียบกับวิธีอื่น เช่น Lesk algorithm ในการหาความคล้ายคลึงกันระหว่างบริบทสั้น ๆ ก็พบว่า LSA ให้ผลดีกว่า (Pino and Eskenazi, 2009) นอกจากนี้ในงานของ Pongpinigpinyo and Rivepiboon ยังได้กล่าวถึงจำนวนของความหมายของคำด้วยว่ามีผลต่อประสิทธิภาพของระบบ โดยคำที่ใช้ศึกษาในงานวิจัยนี้คือ คำนาม หัว มีจำนวนความหมายทั้งหมด 20 ความหมาย และคำกริยา เกือบ มี 9 ความหมาย ซึ่งผลที่ออกมาพบว่าระบบสามารถระบุความหมายของคำว่า เกือบ ได้ถูกต้อง 75.58% ในขณะที่ หัว ถูกต้อง 71.27% Pongpinigpinyo and Rivepiboon จึงตั้งข้อสังเกตว่าจำนวนของความหมายถ้ายิ่งน้อย ประสิทธิภาพของระบบก็น่าจะดีตามไปด้วย เนื่องจากความต่างกันของแต่ละความหมายจะเห็นได้ง่ายและชัดเจนกว่าคำที่มีจำนวนความหมายมากที่ความหมายแต่ละความหมายจะค่อนข้างใกล้เคียงกัน (Pongpinigpinyo and Rivepiboon, 2005) อย่างไรก็ตามในงานดังกล่าวก็เป็นการตั้งข้อสังเกตเปรียบเทียบคำที่มีหน้าที่ของคำ (part of speech) ต่างกัน เมื่อคำมีหน้าที่ต่างกัน ลักษณะของบริบทที่จะนำมาใช้ในการวิเคราะห์ย่อมต่างกันไปด้วย เมื่อเป็นดังนี้ก็อาจกล่าวได้ว่าประสิทธิภาพของระบบอาจขึ้นอยู่กับหน้าที่ของคำก็

เป็นได้กล่าวคือ การที่ระบบสามารถระบุความหมายของ *เก็บ* ได้ดีกว่า *หั่ว* อาจมีเหตุมาจากการที่ *เก็บ* เป็นคำกริยาก็ได้ ดังนั้นการจะเปรียบเทียบว่าจำนวนของความหมายมีผลต่อประสิทธิภาพของระบบหรือไม่ จึงควรเปรียบเทียบคำที่มีหน้าที่ของคำเดียวกัน ดังนั้นในงานวิจัยนี้จึงจะเปรียบเทียบประสิทธิภาพของ LSA ด้วยว่า LSA จะสามารถแยกความหมายของคำแบบใดได้ดีกว่ากัน ระหว่างคำที่มีความหมายน้อยกับคำที่มีความหมายมาก ในขณะเดียวกัน ก็จะเปรียบเทียบประสิทธิภาพของ LSA ในการแยกความหมายของของคำที่มีหน้าที่ต่างกันคือเป็น คำนามและคำกริยา ที่มีจำนวนความหมายใกล้เคียงกันด้วย

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 วิเคราะห์นัยความหมาย¹ของคำกริยาหลายนัยความหมาย และของคำนามหลายนัยความหมาย ประเภทละ 2 คำโดยใช้คลังข้อมูลภาษา

1.2.2 พัฒนาระบบการแยกนัยความหมายของคำหลายนัยความหมายในข้อ 1 โดยใช้แนวทางการวิเคราะห์ความหมายแอบแฝง (Latent Semantic Analysis)

1.2.3 ประเมินความสามารถในการแยกนัยความหมายของระบบ โดยเปรียบเทียบคำหลายนัยความหมายที่อยู่ในประเภทของคำเดียวกันและต่างกัน

1.2.4 ระบุบริบทที่เหมาะสมในการแยกนัยความหมายของคำที่ใช้ในการศึกษา

1.3 สมมติฐานของการวิจัย

1.3.1 คำกริยามีนัยความหมายหลากหลายมากกว่าคำนาม

1.3.2 ระบบสามารถแยกนัยความหมายของคำที่มีความหมายน้อยได้ถูกต้องมากกว่าคำที่มีนัยความหมายมาก และสามารถแยกนัยความหมายของคำนามได้ถูกต้องมากกว่าคำกริยา

1.3.3 บริบทที่เหมาะสมในการแยกนัยความหมายของคำนามคือบริบททางขวา และบริบทที่เหมาะสมในการแยกนัยความหมายของคำกริยาคือบริบททั้งทางซ้ายและทางขวา

¹ นัยความหมาย หมายถึง ความหมายต่าง ๆ ของคำที่ใช้ในงานวิจัยนี้

1.4 ขอบเขตของการวิจัย

1.4.1 ศึกษาคำนาม 2 คำ ได้แก่ *เสียง* และ *หัว* คำกริยา 2 คำ ได้แก่ *บอก* และ *ติด* โดย *เสียง* และ *บอก* เป็นตัวแทนของคำที่มีความหมายน้อย และ *หัว* และ *ติด* เป็นตัวแทนของคำที่มีความหมายมาก

1.4.2 ศึกษาความหมายของคำจากคลังข้อมูลภาษาไทยแห่งชาติด้วยการดึงข้อมูลออกมาคราวละ 200 ตัวอย่าง และวิเคราะห์ความหมายจนกว่าจะไม่เจอความหมายใหม่ในข้อมูลที่ดึงออกมา 2 ครั้งติดกัน

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 เป็นแนวทางในการพัฒนาระบบอื่น ๆ ที่ใช้แนวทางการวิเคราะห์ความหมายแบบแฝง

1.5.2 เป็นแนวทางในการศึกษาพัฒนาการแก้ปัญหาความกำกวมทางความหมายของคำในภาษาอื่น ๆ

1.5.3 เป็นประโยชน์ในการทำพจนานุกรมภาษาไทยแบบอิงคลังข้อมูล

1.6 วิธีดำเนินการวิจัย

1.6.1 ทบทวนวรรณกรรมที่เกี่ยวข้อง ได้แก่ ทฤษฎีเกี่ยวกับคำหลายนัยความหมาย และวิธีการทางคอมพิวเตอร์ที่ใช้ในการแยกนัยความหมายของคำหลายนัยความหมาย

1.6.2 คัดเลือกคำนามหลายนัยความหมาย 2 คำและคำกริยาหลายนัยความหมาย 2 คำ โดยที่ทั้งคำนามและคำกริยาจะมีคำหนึ่งที่มีจำนวนนัยความหมายมากและอีกคำหนึ่งมีจำนวนนัยความหมายน้อยและทั้งนี้จำนวนนัยความหมายของทั้งคำนามและคำกริยาใกล้เคียงกัน

1.6.3 วิเคราะห์นัยความหมายของคำที่ใช้ศึกษาโดยอิงข้อมูลจากคลังข้อมูลภาษาไทยแห่งชาติโดยดึงข้อมูลออกมาครั้งละ 200 ตัวอย่างเพื่อวิเคราะห์ความหมาย และจะสิ้นสุดการวิเคราะห์เมื่อไม่พบความหมายใหม่ในข้อมูลที่ดึงออกมาสองครั้งติดกัน

1.6.4 เตรียมคลังข้อมูลสำหรับใช้ในการทดสอบให้เข้ากับระบบการวิเคราะห์ความหมายแบบแฝง โดยใช้ข้อมูลจากคลังข้อมูลภาษาไทยแห่งชาติ

1.6.5 ทดสอบระบบการวิเคราะห์ความหมายแบบแฝง

1.6.6 ประเมินผลการทดสอบระบบและเปรียบเทียบประสิทธิภาพของระบบของคำทั้ง 4 คำ

1.6.7 ตีความผลการทดสอบในประเด็นที่เกี่ยวข้องกับจำนวนมากน้อยของนัยความหมายชนิดของคำ และบริบทซ้ายหรือขวาที่สามารถใช้ช่วยในการแยกความหมาย

1.6.8 อภิปรายและสรุปผลการวิจัย

1.7 เครื่องมือที่ใช้ในการวิจัย

1.7.1 โปรแกรมภาษา Perl ของบริษัท Active Perl

1.7.2 โปรแกรม Thaiseg version 2.01 ของภาควิชาภาษาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

1.7.3 โปรแกรม Matlab ของบริษัท MathWorks

บทที่ 2

ทบทวนวรรณกรรม

ในทางอรรถศาสตร์ได้แบ่งประเภทของความหมายออกเป็นความหมายประจำรูปภาษา (sense) และความหมายอ้างอิง (reference) โดยที่ความหมายประจำรูปภาษาคือความหมายที่จะจำกัดอยู่ภายในรูปภาษาเท่านั้น แต่ความหมายอ้างอิงคือความหมายที่มีการเชื่อมโยงความหมายประจำรูปคำกับสิ่งใดสิ่งหนึ่ง เช่น *เสื้อขาว* ในทางภาษา *เสื้อ* คือเครื่องแต่งกายชนิดหนึ่ง สำหรับสวม กายท่อนบน และ *ขาว* หมายถึง สีชนิดหนึ่งเหมือนสำลี ตรงข้ามกับสีดำ และในภาษาไทยคำขยาย หรือคำคุณศัพท์มักจะตามหลังคำที่ถูกขยาย ดังนั้นเราจึงรู้ว่านามวลี *เสื้อขาว* หมายถึง เครื่องแต่งกาย ที่ไม่ใช่กระโปรง ไม่ใช่กางเกง มีสีขาว แต่ถ้าเป็นความหมายอ้างอิง *เสื้อขาว* จะต้องอ้างอิงหรือบ่งชี้ไปที่เสื้อตัวใดตัวหนึ่งเป็นพิเศษ (สุริยา รัตนกุล, 2544) สำหรับในงานวิจัยนี้จะศึกษาเฉพาะความหมาย ประจำรูปภาษาเท่านั้น

2.1 คำหลายความหมาย (polysemy)

ปรากฏการณ์ที่คำหนึ่งคำสามารถมีได้หลายความหมายเป็นปรากฏการณ์ทางธรรมชาติของ ภาษาที่พบได้ทั่วไปในภาษาต่าง ๆ ทั่วโลก กล่าวคือไม่จำเป็นต้องมีคำใช้สำหรับทุกความหมายหรือทุก มโนทัศน์ (concept) เพราะไม่เช่นนั้นคำในภาษาก็จะมีมากมาย และขัดกับหลักการประหยัด (economy) ทางภาษา แต่ในทางภาษา นอกจากปรากฏการณ์คำหนึ่งคำมีได้หลายความหมายแล้ว ก็ ยังมีปรากฏการณ์ที่หนึ่งมโนทัศน์สามารถแทนได้ด้วยคำหลายคำได้ด้วยเช่นกัน เกิดเป็นคำพ้อง ความหมาย (synonym) ขึ้นมา

การอธิบายความหมายหรือลักษณะของคำหลายความหมายในทางภาษาศาสตร์ นักภาษาศาสตร์ส่วนใหญ่มักจะอธิบายควบคู่ไปกับคำพ้องรูป (homonymy) เพื่อให้ผู้ใช้ภาษาได้ เข้าใจและเห็นความแตกต่างของลักษณะของคำทั้งสองชนิดนี้ได้ชัดเจนเนื่องจาก ทั้ง homonymy และ polysemy ต่างถูกจัดให้อยู่ในกลุ่มของคำที่มีความกำกวม (ambiguity) ทางความหมาย (Murphy, 2010) โดยที่มโนทัศน์ homonymy คือ ปรากฏการณ์ที่คำในภาษาบังเอิญเหมือนกันโดยที่ ความหมายของคำไม่ได้มีความสัมพันธ์หรือเกี่ยวข้องกัน ซึ่งคำที่บังเอิญเหมือนกันนั้นจัดว่าเป็นคนละ หน่วยศัพท์ (lexeme) กัน ในขณะที่มโนทัศน์ polysemy คือ ปรากฏการณ์ที่คำ ๆ เดียวหรือหน่วย ศัพท์เดียวมีความหมายหลายความหมาย ซึ่งความหมายต่าง ๆ เหล่านี้มีความเกี่ยวข้องสัมพันธ์กัน (Lyons, 1996; Saeed, 2004; Murphy, 2010)

สิ่งสำคัญที่มักนำมาใช้แยกความแตกต่างระหว่างคำพ้องรูปและคำหลายความหมาย คือ ความสัมพันธ์กันทางความหมาย ถ้าความหมายของคำมีความสัมพันธ์กันก็จะจัดให้เป็นความหมายของคำเดียวกันกลายเป็นคำหลายความหมาย แต่ถ้าความหมายของคำไม่มีความสัมพันธ์กันก็จะจัดให้เป็นคำพ้องรูป ซึ่งเกณฑ์ที่มักนำมาใช้ตัดสินว่าความหมายมีความสัมพันธ์กันหรือไม่ คือ การศึกษาดูประวัติความเป็นมาของคำและการใช้สัญชาตญาณ (intuition) ของเจ้าของภาษา (Ravin and Leacock, 2006) ในการศึกษาดูประวัติความเป็นมาของคำ จะตัดสินให้คำ 2 คำเป็นคำพ้องรูป/พ้องเสียงเมื่อประวัติของคำทั้ง 2 คำมีความเป็นมาต่างกัน โดยถือว่ารูปหรือเสียงที่เหมือนกันนั้นเกิดจากความบังเอิญ เช่น ตัวอย่างคำว่า *bass* ในภาษาอังกฤษที่ Ravin และ Leacock (2006) ยกมา *bass* ที่หมายถึงปลาชนิดหนึ่ง มีที่มาจากภาษา Old English *barse* ในขณะที่ *bass* ที่หมายถึงเสียง มาจากภาษาอิตาลี *basso* ดังนั้น *bass* 2 คำนี้จึงจัดว่าเป็นคำพ้องรูป อย่างไรก็ตาม การใช้วิธีการศึกษาดูประวัติความเป็นมาของคำก็ยังเป็นปัญหาในบางครั้งในการแยกคำพ้องรูปออกจากคำหลายความหมาย เนื่องจากมีคำบางคำเมื่อพิจารณาเฉพาะความหมายแล้วจะรู้สึกว่าจะไม่มีความเกี่ยวข้องกัน แต่เมื่อได้วิเคราะห์ไปถึงประวัติความเป็นมาของคำแล้วจะพบว่ามีความเกี่ยวข้องกัน เช่น ตัวอย่างจาก Falkum (2011) คำว่า *cardinal* มีความหมายว่า ผู้นำทางศาสนาของนิกายโรมันคาทอลิก และยังหมายถึงนกชนิดหนึ่งได้อีกด้วย เมื่อดูจากความหมาย ผู้ใช้ภาษามักจะรู้สึกว่าการที่ทั้งสองอย่างนี้ไม่มีความเกี่ยวข้องกัน แต่เมื่อได้ศึกษาถึงที่มาของคำแล้วจะพบว่าทั้งสองความหมายมีที่มาที่สัมพันธ์กัน เนื่องจากนกชนิดนี้ตัวผู้จะมีสีแดง เป็นสีที่ตรงกับชุดคลุมของบาทหลวง *cardinal* ดังนั้นจึงใช้คำว่า *cardinal* สำหรับนกชนิดนี้ด้วย

อีกวิธีหนึ่งที่น่าสนใจแยกความแตกต่างระหว่างคำพ้องรูปกับคำหลายความหมายคือการใช้ความสามารถทางภาษาของเจ้าของภาษาตัดสินว่าความหมายเหล่านั้นสัมพันธ์กันหรือไม่ หากเจ้าของภาษาพิจารณาแล้วเห็นว่าความหมายเหล่านั้นสัมพันธ์กันก็จะตัดสินให้เป็นคำหลายความหมาย แต่ถ้าไม่สัมพันธ์กันก็จะให้เป็นคำพ้องรูปแทน เช่น คำว่า *ปาก* มีความหมายทั้งที่เป็นอวัยวะหนึ่ง ใช้สำหรับกินอาหาร และอีกความหมายคือทางเข้าถ้ำ เมื่อดูจากความหมายแล้วความหมายทั้งสองไม่น่ามีความสัมพันธ์กัน แต่หากพิจารณาให้ถึงลักษณะของอวัยวะปากและปากถ้ำแล้ว จะเห็นว่าทั้งสองความหมายมีความสัมพันธ์กันอยู่ กล่าวคือ *ปาก* ทั้งสองความหมายต่างเป็นส่วนที่เปิดรับให้สิ่งภายนอกเข้าไปด้านในได้หรือให้สิ่งที่อยู่ภายในออกมาได้เหมือนกัน *ปาก* ที่เป็นอวัยวะก็เป็นทางเข้าของอาหารให้เข้าไปเลี้ยงร่างกาย หรือเอาสิ่งที่อยู่ภายในออกมา เช่น การคาย การอ้วก เป็นต้น ซึ่งเหมือนกับ *ปาก* ที่หมายถึง “ปากถ้ำ” อันเป็นส่วนที่เป็นทางเข้าออกให้สิ่งภายนอกเข้าไปในถ้ำได้ หรือให้สิ่งที่อยู่ภายในออกมาได้เช่นกัน อย่างไรก็ตามการใช้ความสามารถของเจ้าของภาษาในการตัดสินก็ยังมีจุดบกพร่องอยู่ เพราะการตัดสินด้วยวิธีนี้จะมีลักษณะเป็นลำดับขั้น (degree) ที่จะให้

ความหมายมีความสัมพันธ์กันน้อยเพียงใด เจ้าของภาษาคนหนึ่งอาจตัดสินว่าความหมายที่ให้มาพิจารณามีความสัมพันธ์กันมาก แต่เจ้าของภาษาอีกคนหนึ่งอาจเห็นต่างออกไป เห็นว่ามีความสัมพันธ์กันน้อย แบบนี้แล้วก็ยากจะตัดสินได้ว่าจะกำหนดให้เป็นคำหลายความหมายหรือคำพ้องรูป เช่น ในภาษาอังกฤษ คำว่า *ear* ที่หมายถึงอวัยวะหนึ่ง ทำหน้าที่ฟังเสียง กับอีกความหมายหนึ่งคือ ส่วนที่เป็นรวงของพืช เช่น *an ear of corn* หากถามเจ้าของภาษาถึงความหมายทั้งสองว่ามีความสัมพันธ์กันหรือไม่ ก็จะมีทั้งกลุ่มที่เห็นว่าความหมายทั้งสองสัมพันธ์กันกับอีกกลุ่มที่เห็นว่าความหมายไม่มีความสัมพันธ์กันเลย (Falkum, 2011) เป็นต้น

จากหลักเกณฑ์ทั้งสองหลักเกณฑ์ที่กล่าวไปจะเห็นว่ายังเป็นเกณฑ์ที่ไม่สามารถนำมาใช้ตัดสินคำศัพท์ได้ทุกคำว่าควรเป็นคำหลายความหมายหรือคำพ้องรูป สิ่งที่แสดงให้เห็นได้ชัดเจนว่าหลักเกณฑ์ยังมีข้อบกพร่องอยู่ คือพจนานุกรมคำศัพท์ ในพจนานุกรม คำที่เป็นคำพ้องรูปจะแยกไว้เป็นคนละคำศัพท์ แต่ถ้าเป็นคำหลายความหมาย ความหมายต่าง ๆ จะรวมไว้อยู่ในคำศัพท์เดียว เมื่อลองเปรียบเทียบพจนานุกรมคำศัพท์ต่าง ๆ แล้วจะพบว่าจำนวนความหมายของคำบางคำมีไม่เท่ากัน นั่นหมายความว่าเกณฑ์การแบ่งความหมายของนักพจนานุกรมแต่ละคนไม่เหมือนกัน ซึ่ง Fillmore และ Atkins (2006) ได้ให้ข้อสังเกตว่าการจะจับคู่ความหมายของคำจากพจนานุกรมหนึ่งไปหาอีกพจนานุกรมนั้นแทบเป็นไปไม่ได้เลย แม้แต่กับคำที่มีความหมายกำกวมไม่มากก็ตาม

2.2 ทฤษฎีหรือแนวคิดที่เกี่ยวข้องกับคำหลายความหมาย

ทฤษฎีสำคัญที่เกี่ยวข้องกับคำหลายความหมาย ได้แก่ ทฤษฎีต้นแบบ และทฤษฎีคำหลายความหมายอย่างมีหลักการ (principled polysemy) ทฤษฎีต้นแบบมีแนวคิดที่สำคัญ คือความหมายที่เป็นศูนย์รวมของความหมายอื่น ๆ จะถือว่าเป็นความหมายต้นแบบของคำนั้น ๆ และความหมายอื่น ๆ ก็ปรับเปลี่ยนคุณสมบัติบางอย่างจากความหมายต้นแบบทำให้เกิดเป็นคำหลายความหมาย อย่างไรก็ตามทฤษฎีนี้ยังมีข้อบกพร่องที่ไม่มีเกณฑ์ที่แน่นอนในการหาความหมายของคำ รวมถึงใช้ความเห็นของผู้วิเคราะห์ในการแยกความหมายของคำมากเกินไป ดังนั้นทฤษฎีคำหลายความหมายอย่างมีหลักการจึงเกิดขึ้นเพื่อแก้ไขจุดบกพร่องดังกล่าว

2.2.1 ทฤษฎีต้นแบบ (prototype theory)

ทฤษฎีนี้ Rosch ได้พัฒนาขึ้นในช่วงยุคปีค.ศ.1970 จากการทดลองของ Rosch (1977 อ้างถึงใน Ravin and Leacock, 2006) พบว่าคนเราจะจัดกลุ่มวัตถุโดยการนำเอาวัตถุนั้นไปเทียบกับสิ่งที่เป็นต้นแบบของกลุ่มนั้น ๆ ซึ่งลักษณะของการจัดกลุ่มนั้นจะคล้ายกับแนวคิดเรื่องระดับชั้น (hierarchy) เช่น สุนัขเป็นสัตว์เลี้ยงลูกด้วยนม เป็นสัตว์ และเป็นสิ่งมีชีวิต จะเห็นว่ากลุ่มที่สุนัขอยู่มีลักษณะเรียงลำดับจากกลุ่มเล็กไปหากลุ่มใหญ่ แต่ถ้าเป็นระดับชั้นทางมโนทัศน์ (conceptual

category) จะไม่มีการแบ่งแยกเป็นกลุ่มจากกันชัดเจน แต่จะมีลักษณะเป็นระดับ (degree) ความใกล้เคียงจากลักษณะต้นแบบ Rosch ได้เสนอโมเดลลักษณะต้นแบบไว้ 2 โมเดล คือ แบบที่มีต้นแบบเพียงแบบเดียว เป็นศูนย์รวมของคุณสมบัติ (characteristic feature) ต่าง ๆ กับแบบที่มีต้นแบบหลายต้นแบบ และแต่ละต้นแบบจะมีชุดของคุณสมบัติต่าง ๆ แตกต่างกันไป โดยที่คุณสมบัตินี้ไม่จำเป็นต้องเหมือนกัน ซึ่งรูปแบบหลังนี้นักภาษาศาสตร์ได้นำมาปรับใช้กับการวิเคราะห์คำหลายความหมาย

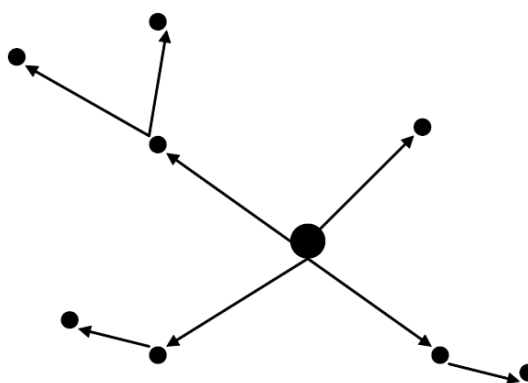
George Lakoff (Evans and Green, 2006) กล่าวว่าคำศัพท์ (lexical item) ต่าง ๆ มีลักษณะคล้ายกับกลุ่มของรัศมีที่แผ่กระจาย (radial categories) ซึ่งรัศมีที่แผ่กระจายออกมานี้ก็เปรียบเหมือนกับเป็นความหมายของคำนั่นเอง Lakoff มองว่าความหมายที่เจ้าของภาษาส่วนใหญ่ตัดสินว่าเป็นความหมายที่ชัดเจนกว่าความหมายอื่น ถือว่าเป็นความหมายที่มีลักษณะใกล้เคียงกับต้นแบบมากที่สุด ซึ่งความหมายนี้จะจัดว่าเป็นความหมายศูนย์กลาง (central sense) แต่ถ้าเป็นความหมายที่ใกล้เคียงกับต้นแบบนี้ลดลง ก็จะอยู่ห่างออกจากศูนย์กลางออกไป เปรียบเหมือนกับอยู่ในช่วงรัศมีที่แผ่ออกมา จัดว่าเป็นความหมายชายขอบ (peripheral sense) เช่น ตัวอย่างคำบุพบท *over* ในภาษาอังกฤษ

- | | | |
|-----|---|-------------------|
| (1) | a. The picture is <u>over</u> the sofa. | ABOVE |
| | b. The ball landed <u>over</u> the wall. | ON THE OTHER SIDE |
| | c. The car drove <u>over</u> the bridge. | ACROSS |
| (2) | Jane has a strange power <u>over</u> him. | CONTROL |

จากตัวอย่างในข้อ (1) ความหมายของ a. b. และ c. จะต่างกันเพียงเล็กน้อย และตัวอย่าง (1a) เป็นตัวอย่างที่เจ้าของภาษาส่วนใหญ่มีความเห็นตรงกันว่าเป็นตัวอย่างที่แสดงถึงความหมายของคำว่า *over* ได้ดีกว่าตัวอย่างอื่น ๆ ดังนั้นความหมาย (1a) จึงจะเป็นความหมายที่อยู่ใกล้เคียงกับลักษณะต้นแบบมากที่สุด ส่วนความหมาย (2) หากพิจารณาโดยผิวเผินแล้วจะเห็นว่าต่างจากความหมาย (1) ซึ่งหากวิเคราะห์ให้ความหมาย (2) ไม่เกี่ยวข้องกับความหมายใน (1) เลย คำว่า *over* (1) และ (2) ก็จะจัดว่าเป็นคำพ้องรูปกัน แต่หากวิเคราะห์ว่าความหมายสามารถแบ่งออกได้เป็นความหมายตรงตัว (spatial sense) ซึ่งความหมายแบบตรงตัวนี้นักภาษานุกรมมักจะจัดให้เป็นความหมายหลัก (primary sense) ของคำศัพท์นั้น ๆ กับความหมายโดยนัย (non-spatial sense) และความหมาย (2) เป็นความหมายโดยนัยที่เกิดจากกระบวนการอุปลักษณ์ (metaphor) เช่นนี้ความหมาย (1) และ (2) ก็จัดว่าเป็นความหมายของคำศัพท์เดียวกันซึ่งเป็นคำหลายความหมาย จากตัวอย่างการวิเคราะห์จะเห็นว่าเราจะวิเคราะห์ให้คำศัพท์เป็นคำพ้องรูปหรือคำหลายความหมายนั้น

ขึ้นอยู่กับเหตุผลของผู้วิเคราะห์ ซึ่งทำให้เห็นว่าการตัดสินใจให้คำศัพท์เป็นคำพ้องรูปหรือคำหลายความหมายนั้นยังเป็นปัญหาอยู่

ตามแนวคิดของ Lakoff ความหมายต่าง ๆ ของคำหลายความหมายจัดว่าเป็นกลุ่มของความหมายหนึ่งกลุ่ม (single category) โดยที่ความหมายใกล้เคียงกับความหมายต้นแบบจะอยู่ใกล้ศูนย์กลางทางความหมายและความหมายชายขอบจะอยู่ไกลจากความหมายศูนย์กลางออกไป แผนภาพแสดงลักษณะการขยายทางความหมายของคำหลายความหมายตามแนวคิดของ Lakoff สามารถดูได้จากภาพที่ 2.1



ภาพที่ 2.1 แผนภาพแสดงการขยายทางความหมายของ radial categories (Evans and Green, 2006)

จากภาพ วงกลมสีดำแทนความหมายโดยที่วงกลมตรงกลางที่ใหญ่กว่าวงกลมอื่นแทนความหมายต้นแบบและลูกศรแทนทิศทางการขยายทางความหมาย จะเห็นว่าความหมายชายขอบที่อยู่ไกลออกไปจากศูนย์กลาง เป็นความหมายที่พัฒนามาจากความหมายที่อยู่ใกล้ศูนย์กลางมากกว่า ซึ่งนั่นเกิดจากกลไกการขยายของความหมาย (meaning extension) ทั้งนี้ความหมายที่ขยายออกไปยังรวมถึงความหมายเชิงอุปลักษณ์ด้วย

อย่างไรก็ตาม การวิเคราะห์ความหมายของ Lakoff ยังมีข้อบกพร่องอยู่ เนื่องจากการวิเคราะห์ที่ไม่มีลักษณะเป็นการให้ความหมายแบบละเอียดมาก (full-specification approach) ดังนั้นความหมายของคำหนึ่งคำจึงมีมากไปด้วย เนื่องจากนักภาษาศาสตร์ปริชาณมักไม่สนใจประเด็นเรื่องความประหยัด (economy) (Evans and Green, 2006) จุดบกพร่องของการวิเคราะห์คำหลายความหมายของ Lakoff มี 2 ประเด็นใหญ่ ๆ ด้วยกัน คือ

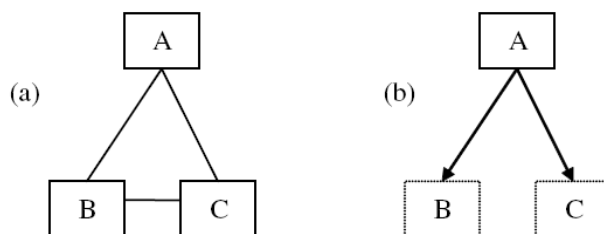
1. ไม่มีการแยกความแตกต่างระหว่างคำหลายความหมายกับคำที่ไม่ได้มีหลายความหมาย แต่ความหมายแตกต่างกันไปเพราะบริบทที่คำ ๆ นั้นไปปรากฏอยู่หรือเรียกว่าคำที่มีความหมายคลุมเครือ (vagueness) เช่น คำว่า *thing* ที่สามารถหมายถึงสิ่งใดก็ได้ และเราจะรู้ความหมายว่า *thing* หมายถึงสิ่งใด เมื่อ *thing* ไปอยู่ในบริบทต่าง ๆ แล้ว อีกตัวอย่างหนึ่งคือ *aunt* ซึ่งสามารถ

หมายถึงป่าทางฝ่ายพ่อหรือฝ่ายแม่ก็ได้ โดยที่เรารู้ได้ว่าเป็นป่าจากทางฝ่ายใดก็ต้องอาศัยบริบทที่ *aunt* ไปปรากฏอยู่ แต่ถ้าหากเป็นความหมายของคำหลายความหมายแล้วจะใช้ความหมายประจำคำได้เลย โดยไม่จำเป็นต้องอาศัยบริบทช่วย ตัวอย่างที่แสดงให้เห็นว่า Lakoff ไม่ได้นำประเด็นนี้มาเป็นเกณฑ์ในการวิเคราะห์ความหมายคือความหมายของคำว่า *over* ที่แยกความหมายเรื่องการสัมผัสกัน (contact) ระหว่างผู้กระทำกับกรรม (Evans and Green, 2006) เช่น

(3) a. The bird flew over the wall.

b. Sam climbed over the wall.

จากตัวอย่าง Lakoff จะวิเคราะห์ให้ความหมายของ *over* ใน (3a) และ (3b) เป็นคนละความหมายกัน โดยมีความแตกต่างกันในเรื่องการสัมผัส ซึ่งการวิเคราะห์เช่นนี้ Tyler และ Evans (Tyler and Evans, 2003 อ้างถึงใน Evans and Green, 2006) มองว่าในความเป็นจริงเรารู้ได้ว่าผู้กระทำกับกรรมจะมีการสัมผัสกันหรือไม่ เราต้องอาศัยความรู้ทางโลกของเราด้วย เช่น เราทราบว่าบินได้ดังนั้นนกจึงไม่สัมผัสถูกกำแพง ในขณะที่คนบินไม่ได้ดังนั้นในการปีนกำแพงก็ต้องมีการสัมผัสกันเป็นธรรมดา ดังนั้นความแตกต่างกันในเรื่องการสัมผัสนี้จึงไม่สมควรนำมาใช้แยกความหมายของ *over* ให้เป็นคนละความหมายกัน ภาพแสดงความแตกต่างระหว่างคำหลายความหมายกับคำที่คลุมเครือต้องอาศัยบริบทได้แสดงไว้ในภาพที่ 2.2



ภาพที่ 2.2 แผนภาพแสดงความแตกต่างระหว่างความหมายของคำหลายความหมาย กับความหมายที่ต้องอาศัยบริบท (Evans and Green, 2006)

ภาพ (a) แสดงถึงความหมายของคำหลายความหมายที่แต่ละความหมายอยู่แยกกันชัดเจน แสดงเป็นเส้นทึบและเส้นที่โยงกันไม่มีหัวลูกศรเพราะแต่ละความหมายมีความสัมพันธ์กัน แต่ภาพ (b) แสดงถึงความหมายที่ต้องอาศัยบริบทจะเห็นได้ว่าความหมายของคำคือ A และเมื่อไปอยู่ในบริบทที่แตกต่างกันไปก็จะได้ความหมาย B และ C ซึ่งความหมาย B และ C เป็นความหมายที่ได้จากการตีความผ่านบริบท แสดงด้วยสัญลักษณ์เส้นประ

2. เกณฑ์ที่ใช้ในการวิเคราะห์ความหมายของคำหลายความหมายไม่ชัดเจน เนื่องจาก Lakoff ไม่ได้กำหนดเกณฑ์สำหรับแยกความต่างระหว่างความหมายออกจากกัน ทำให้ไม่รู้ว่เมื่อใดจึง

จะควรแยกหรือไม่ควรแยกความหมาย ซึ่งนั่นหมายความว่า Lakoff ใช้ความรู้ทางภาษาของเจ้าของภาษาในการวิเคราะห์เท่านั้น ทำให้แนวคิดของ Lakoff ยากจะนับให้เป็นทฤษฎีทางภาษาศาสตร์หนึ่ง เนื่องจากขาดหลักการที่มีลักษณะเป็นวิทยาศาสตร์เป็นเหตุเป็นผลกัน (Sandra, 1998 อ้างถึงใน Evans and Green, 2006)

จากข้อบกพร่องดังกล่าว Vyvyan Evans และ Andrea Tyler จึงพัฒนาทฤษฎีใหม่ขึ้นมา ชื่อ ทฤษฎีคำหลายความหมายอย่างมีหลักการ (principled polysemy)

2.2.2 ทฤษฎีคำหลายความหมายอย่างมีหลักการ (principled polysemy)

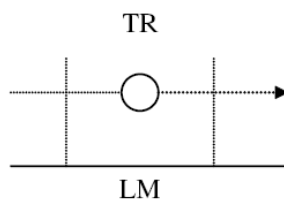
ทฤษฎีนี้ Vyvyan Evans และ Andrea Tyler พัฒนาขึ้นเพื่อหาหลักการที่มีลักษณะเป็นอวัตถุวิสัย (objectivity) ในการแยกความหมายของคำหลายความหมาย เนื่องจากแนวคิดของ Lakoff ที่ผ่านมาถูกวิจารณ์ว่าไม่มีหลักเกณฑ์ที่แน่นอนในการแยกความหมาย และใช้ความเห็นของผู้วิเคราะห์เป็นหลักในการหาความหมายและแยกความหมายของคำ มีลักษณะเป็นอัตวิสัย (subjectivity) มากเกินไป สำหรับทฤษฎีคำหลายความหมายอย่างมีหลักการนี้พยายามหาหลักเกณฑ์เพื่อให้สามารถแยกความหมาย (Evans and Tyler, 2003; Evans and Green, 2006) ต่อไปนี้ได้

1. หาความหมายที่เป็นต้นแบบหรือความหมายศูนย์กลาง (central sense) ของกลุ่มความหมาย (radical category) ได้

2. แยกความหมายต่าง ๆ ของคำหลายความหมายได้ รวมถึงแยกความหมายที่ไม่ต้องอาศัยบริบทออกจากความหมายที่ต้องอาศัยบริบท (context-dependent) ได้ หรือกล่าวได้อีกนัยหนึ่งว่าเป็นการแยกระหว่างความหมายของคำจริง ๆ ออกจากความหมายที่ไม่ชัดเจน คลุมเครือ (vagueness)

สำหรับเกณฑ์การแยกความหมายของคำบุพบท เช่น คำว่า *over* ที่ Lakoff ศึกษา นั้น Tyler และ Evans ได้กำหนดเกณฑ์ขึ้นมา 2 ข้อ คือ

1. ความหมายนั้นจะต้องไม่มีความสัมพันธ์ใด ๆ (non-spatial) กับส่วนที่เป็น TR (trajector) ซึ่งเป็นส่วนที่เคลื่อนที่ได้ และ LM (landmark) ซึ่งเป็นส่วนที่เคลื่อนที่ไม่ได้ TR และ LM นี้เป็นสิ่งที่ Lakoff ตั้งขึ้นเพื่อแสดงแผนภาพการใช้คำบุพบท เป็นแผนภาพแสดงความสัมพันธ์ระหว่าง TR กับ LM เช่น ประโยค *The bird flew over the yard.* จะได้แผนภาพแสดงความสัมพันธ์ระหว่าง TR และ LM ดังนี้



ภาพที่ 2.3 The bird flew over the yard. ปรับจากงานของ Lakoff (Evans and Green, 2006)

จากภาพ *bird* เป็น TR และ *yard* เป็น LM *over* แสดงความสัมพันธ์ระหว่าง *bird* กับ *yard* ว่า *bird* อยู่ตำแหน่งสูงกว่า *yard* เป็นต้น

2. ความหมายนั้นต้องไม่พึ่งพิงบริบท (context-independent)

ทั้งสองเกณฑ์นี้ Tyler และ Evans นำมาใช้วิเคราะห์คำบุพบทในภาษาอังกฤษ นอกเหนือจากคำบุพบท Evans (2004 อ้างถึงใน Evans and Green, 2006) ได้พัฒนาเกณฑ์สำหรับ วิเคราะห์คำหลายความหมายประเภทอื่นด้วย เช่น คำนาม คำกริยา โดยเกณฑ์ที่ใช้มี 3 เกณฑ์ด้วยกัน คือ

1. เกณฑ์ทางความหมาย (meaning criterion) คือ ความหมายที่จะแยกออกไปเป็นอีก ความหมายได้จะต้องมีความหมายใหม่เพิ่มเข้ามา ซึ่งไม่สามารถพบได้ในความหมายอื่น

2. เกณฑ์ด้านการอธิบายมโนทัศน์ (concept elaboration criterion) คือ ความหมายที่ ต่างกันจะมีรายละเอียดรูปแบบของมโนทัศน์ที่ต่างกัน ซึ่งสามารถแสดงให้เห็นได้ในระดับภาษา (linguistic level) เช่น แสดงออกมาในรูปของคำที่ปรากฏรวมว่าควรใช้คำใดเพื่อช่วยแสดงให้เห็นมโน ทัศน์นั้น เช่น การเลือกใช้คำคุณศัพท์ขยายคำนาม หากคำนามนั้นแสดงมโนทัศน์ที่ต่างกัน คำคุณศัพท์ ที่จะนำมาใช้ขยายก็จะต่างกันไปด้วย

3. เกณฑ์ทางไวยากรณ์ (grammatical criterion) คือ ความหมายที่ต่างกันจะมีรูปแบบ การพึ่งพาทางโครงสร้าง (structural dependency) ของภาษาที่ต่างกัน กล่าวคือ เมื่อความหมาย ต่างกันลักษณะโครงสร้างทางไวยากรณ์ที่จะใช้สำหรับความหมายนั้น ๆ ก็อาจจะมีลักษณะเฉพาะที่ ต่างกันไปด้วย

Evans แสดงตัวอย่างการวิเคราะห์ความหมายของคำนาม *time* โดยใช้เกณฑ์ทั้ง 3 ข้อนี้นี้ใน การวิเคราะห์ตัวอย่างประโยคต่อไปนี้

(4) Time flies when you're having fun.

(5) The time has arrived to finally tackle environmental pollution.

ในเกณฑ์ด้านความหมาย จากตัวอย่าง (4) *time* ให้ความหมายถึงช่วงเวลาหรือ DURATION ว่าช่วงเวลาที่มีความสุขจะผ่านไปอย่างรวดเร็ว แต่ในตัวอย่าง (5) ความหมายจะต่างออกไป *time* ไม่ได้แสดงช่วงเวลาแต่แสดงถึงจุดใดจุดหนึ่งของเวลา หรือ MOMENT จากเกณฑ์ด้านความหมายจะเห็นว่าความหมายทั้งสองต่างกัน เมื่อมาวิเคราะห์โดยใช้เกณฑ์การอธิบายโน้ตส์ ตัวอย่างทั้งสองมีโน้ตส์ที่ต่างกัน โดยทั้งสองเกี่ยวข้องกับโน้ตส์การเคลื่อนไหว (motion) แต่ในตัวอย่าง (4) จะแสดงถึงโน้ตส์การเคลื่อนไหวในด้านความเร็ว แต่ตัวอย่าง (5) จะเป็นการเคลื่อนไหวที่เป็นลักษณะของช่วงเวลาจริง ๆ มีจุดอ้างถึงที่เฉพาะเจาะจงว่าเป็นเวลาใดได้ เช่น ขณะนี้ (NOW) และเมื่อมาพิจารณาถึงเกณฑ์ทางไวยากรณ์ โดยปกติ *time* จะเป็นคำนามนับไม่ได้ ดังนั้นจึงจะใช้คำนำหน้านามเป็นเอกพจน์ไม่ได้ แต่ถ้าเป็นกรณีที่ *time* หมายถึง MOMENT จะถือว่าเป็นคำนามนับได้ และสามารถใช้นำหน้านามเอกพจน์ได้ ดังตัวอย่าง (6)-(7)

(6) *A time raced by.

(7) A time will come when we'll finally have to address global warming.

จากการวิเคราะห์คำว่า *time* โดยใช้เกณฑ์ทั้ง 3 เกณฑ์จะเห็นว่า *time* ที่หมายถึง DURATION และ MOMENT ถือได้ว่าเป็นคนละความหมายกัน

สำหรับงานวิจัยนี้จะใช้หลักการวิเคราะห์ความหมายของคำตามแนวทางของทฤษฎีคำหลายความหมายอย่างมีหลักการ เนื่องจากมีลักษณะการวิเคราะห์แบบภววิสัยและมีเกณฑ์การวิเคราะห์ชัดเจนกว่า

2.3 การศึกษาความหมายในทางภาษาศาสตร์คอมพิวเตอร์

ความสนใจเรื่องความหมายของคำในทางคอมพิวเตอร์มีมานานแล้วตั้งแต่ช่วงปลายยุคปี ค.ศ.1940 (Agirre and Edmons, 2007) โดยเน้นไปที่การศึกษาเรื่องความกำกวมทางความหมายของคำ (Word Sense Disambiguation: WSD) ซึ่งความกำกวมที่กล่าวถึงนี้ยังรวมถึงความหมายของคำหลายความหมายด้วย เป็นการสอนให้คอมพิวเตอร์สามารถแยกความหมายของคำเมื่ออยู่ในบริบทต่าง ๆ ซึ่งการศึกษาในเรื่องความหมายนี้ก็เพื่อนำไปเป็นส่วนหนึ่งในการพัฒนาระบบการแปลภาษาด้วยเครื่อง (Machine Translation: MT) ซึ่งการศึกษาในช่วงแรกจะใช้หลักสถิติในการช่วยหาความหมาย อย่างไรก็ตามการแยกความหมายของคำก็ถือเป็นเรื่องยาก ทำให้การศึกษาคำความหมายของคำและการแปลภาษาด้วยเครื่องลดน้อยลงในช่วงยุคปีค.ศ.1960 จนถึงช่วงปีค.ศ.1980 ถือเป็นจุดเปลี่ยนของการศึกษาคำความกำกวมทางความหมายของคำ เนื่องจากเป็นยุคที่เริ่มมีการสร้างคลังข้อมูล การวิเคราะห์สิ่งต่าง ๆ ด้วยมือเริ่มถูกแทนที่ด้วยการสกัดข้อมูลต่าง ๆ ออกมาจากคลังข้อมูลแบบอัตโนมัติ ปัจจุบันวิธีการให้เครื่องเรียนรู้จากคลังข้อมูลที่มีการกำกับความหมายของคำ

ไว้ถือเป็นแนวทางกระแสหลักในการศึกษาความกำกวมทางความหมายของคำ (Agirre and Edmons, 2007)

ในหัวข้อถัดไปผู้วิจัยจะกล่าวถึงแนวทางหลัก ๆ ที่ใช้การศึกษาความกำกวมทางความหมายของคำในทางภาษาศาสตร์คอมพิวเตอร์ว่ามีแนวทางใดบ้าง แล้วจากนั้นจะอธิบายถึงวิธี Latent Semantic Analysis (LSA) ซึ่งเป็นวิธีการที่จะนำมาใช้ในงานวิจัยชิ้นนี้ว่ามีหลักในการทำงานอย่างไรเป็นหัวข้อถัดไป

2.4 แนวทางการศึกษาความกำกวมทางความหมายของคำ

แนวทางการศึกษาความกำกวมทางความหมายของคำสามารถแบ่งออกเป็นวิธีการหลัก ๆ ได้ 3 วิธี (Navigli, 2009; Niruddin, 2013) ดังนี้

1. วิธีให้เครื่องเรียนรู้แบบมีผู้สอน (Supervised WSD)

วิธีการนี้จะให้คอมพิวเตอร์เรียนรู้จากคลังข้อมูลสำหรับฝึกฝน (training corpus) ที่มีการกำกับข้อมูลต่าง ๆ ไว้ให้แล้ว เช่น ความหมายของคำเป็นต้น ส่วนใหญ่มักต้องใช้แรงงานคนหรือให้ผู้เชี่ยวชาญด้านภาษาเป็นผู้กำกับข้อมูลลงไป วิธีการนี้เป็นวิธีที่ได้ผลมากที่สุดในงานด้าน word sense disambiguation เมื่อเทียบกับวิธีอื่น รวมถึงเป็นวิธีที่นิยมใช้กัน (Edmonds, 2006; Agirre and Edmons, 2007) วิธีการให้เครื่องเรียนรู้มีเทคนิคที่สำคัญคือให้เครื่องเรียนรู้จากคลังข้อมูลสำหรับฝึกฝนที่จะมีตัวอย่างข้อมูลของคำเป้าหมาย (target word) อยู่ ซึ่งคลังข้อมูลที่ใช้ฝึกฝนจะต้องเป็นคลังข้อมูลขนาดใหญ่ เพราะขนาดของคลังข้อมูลมีผลต่อประสิทธิภาพของระบบ เนื่องจากถ้าคลังข้อมูลมีขนาดเล็ก ขนาดตัวอย่างที่จะให้เครื่องเรียนรู้ก็จะน้อยตามไปด้วยไม่เพียงพอต่อการเรียนรู้ ดังนั้นหากเลือกใช้วิธีนี้ ยิ่งขนาดคลังข้อมูลมีมากเท่าไรก็มีแนวโน้มว่าประสิทธิภาพของระบบก็จะดีขึ้นตามไปด้วย นอกจากคลังข้อมูลแบบภาษาเดียว ยังอาจใช้คลังข้อมูลแบบสองภาษา ก็ได้ (parallel corpora) เป็นคลังข้อมูลที่เป็นลักษณะเหมือนการแปลสองภาษา คือ ประโยคในภาษาที่หนึ่งจะจับคู่กับประโยคในภาษาที่สอง เพื่อให้สามารถรู้ได้ว่าคำในภาษาที่หนึ่งตรงกับคำว่าอะไรในภาษาที่สอง เนื่องจากมีความเชื่อว่าคำหลายความหมายในภาษาหนึ่งเมื่อแปลไปเป็นภาษาอื่นแล้ว แต่ละความหมายจะใช้คำแปลคนละคำกัน เช่น คำว่า *partido* ในภาษาสเปน หากเป็นความหมายที่เกี่ยวกับกีฬา จะใช้คำว่า *match* ในภาษาอังกฤษ แต่ถ้าเป็นความหมายในเชิงการเมืองจะใช้คำว่า *party* แทน (Màrquez et al., 2007)

สิ่งสำคัญสำหรับวิธีการให้เครื่องเรียนรู้คือ คุณสมบัติ (feature) ที่กำหนดให้กับระบบ เพราะคุณสมบัติจะเป็นข้อมูลต่าง ๆ ของบริบทรอบ ๆ คำเป้าหมายที่จะช่วยให้ระบบสามารถแยกความหมายของคำได้ คุณสมบัติที่ใช้สามารถแบ่งออกได้เป็น 3 กลุ่ม (Màrquez et al., 2007) ดังนี้

1. Local feature เป็นคุณสมบัติที่ให้ข้อมูลเกี่ยวกับคำปรากฏร่วมของคำเป้าหมาย เช่น n-gram ของชนิดของคำ รูปคำปกติ (lemma) หรือรูปคำที่ปรากฏในบริบท (word form) และตำแหน่งต่าง ๆ ของคำที่อยู่ใกล้กับคำเป้าหมาย เป็นต้น

2. Topical feature คุณสมบัตินี้จะให้ข้อมูลเกี่ยวกับขอบเขตความหมาย (semantic domain) ของข้อความหรือเอกสาร ขนาดของบริบทที่ใช้ในคุณสมบัตินี้จะกว้างกว่าของ Local feature อาจจะใช้เป็นย่อหน้าหรือข้อความทั้งเอกสาร คุณสมบัติเช่น bag-of-words เป็นการแสดงความถี่ของคำว่าแต่ละคำในเอกสารมีจำนวนเท่าใด โดยไม่สนใจลำดับทางไวยากรณ์ คำใดที่มีจำนวนมาก เมื่อนำมาประมวลก็จะทำให้รู้ได้ว่าเอกสารนั้นเกี่ยวข้องกับเรื่องอะไร

3. Syntactic dependency เป็นคุณสมบัติที่อยู่ในระดับประโยค แสดงถึงความสัมพันธ์ทางไวยากรณ์ของคำต่าง ๆ

อัลกอริทึมหรือโมเดลที่ใช้มีหลายวิธีด้วยกัน เช่น Decision Tree, Decision List, Naïve Bayes, Support Vector Machines (SVM), Maximum Entropy (ME) เป็นต้น ซึ่งโมเดลที่เป็นที่นิยมใช้คือ SVM เนื่องจากเป็นวิธีที่ให้ค่าความถูกต้องได้ดีกว่าวิธีอื่น (Edmonds, 2006; Niruddin, 2013) เนื่องจาก SVM สามารถจัดการกับขนาดของคุณสมบัติที่ใหญ่และซับซ้อน (very high-dimensional feature space) ได้ดี

นอกจากนี้ยังมีวิธีที่ให้เครื่องเรียนรู้แบบมีผู้สอน (Minimally-supervised WSD) คือให้เครื่องเรียนรู้จากข้อมูลบางส่วนที่มีการกำกับข้อมูลความหมายของคำไว้ซึ่งเป็นข้อมูลไม่มากเท่ากับการใช้วิธีให้เครื่องเรียนรู้แบบมีผู้สอนแล้วจากนั้นให้เครื่องนำเอาความรู้ที่ได้จากการเรียนรู้ข้อมูลนั้น มาสกัดเอาความหมายออกมาจากคำที่อยู่ในคลังข้อมูลดิบที่ไม่มีข้อมูลใดกำกับไว้เลย

แม้วิธีการให้เครื่องเรียนรู้แบบมีผู้สอนจะให้ผลดี แต่ก็มีข้อจำกัดว่าต้องใช้คลังข้อมูลที่มีการกำกับความหมายของคำไว้แล้วในคลังข้อมูลสำหรับฝึกฝน ซึ่งต้องใช้แรงงานคนในการกำกับความหมายและผู้ที่กำกับต้องเป็นผู้มีความรู้ทางด้านภาษาและความหมายดีพอสมควร เช่น นักทำพจนานุกรมหรือนักภาษาศาสตร์ ดังนั้นการสร้างคลังข้อมูลจึงใช้เวลาค่อนข้างมาก ปัจจุบันมีคลังข้อมูลที่มีการกำกับความหมายของคำและมักนำมาใช้ในงานวิจัยเกี่ยวกับการแก้ปัญหาคำกำวมของคำ เช่น Semcor เป็นข้อมูลส่วนหนึ่งจาก Brown Corpus หรือ DSO corpus ซึ่งเป็นข้อมูลประโยคจาก Wall Street Journal corpus (WSJ) และ Brown Corpus เป็นต้น จะเห็นว่าคลังข้อมูลที่มีจะเป็นภาษาอังกฤษ ซึ่งถ้าเป็นภาษาอื่นที่ขาดแคลนคลังข้อมูลเช่นนี้ก็ยากที่จะใช้วิธีการให้เครื่องเรียนรู้แบบมีผู้สอนได้ เพราะระบบต้องมีการฝึกฝนก่อนจึงจะสามารถแยกความกำวมของคำได้ อีกทั้งถึงแม้คลังข้อมูลที่ใช้ศึกษาในภาษาอังกฤษจะมีแต่ถ้ามีคำใหม่หรือคำยืมเกิดขึ้น

ความหมายที่กำกับไว้แล้วก็จะไม่ครอบคลุมความหมายทั้งหมดที่ใช้จริง นอกจากนี้การจะนำเอาระบบที่ใช้ในภาษาหนึ่งได้ดีไปปรับใช้กับอีกภาษาหนึ่งก็เป็นไปได้ยาก เพราะโครงสร้างทางภาษาไม่เหมือนกัน อีกทั้งผลที่ได้ไม่แน่นอน กล่าวคือ เมื่อนำไปใช้กับภาษาหนึ่งแล้วได้ผลดี ไม่ได้หมายความว่า จะนำไปใช้กับอีกภาษาหนึ่งแล้วจะได้ผลดีตามไปด้วยเสมอไป

2. วิธีการใช้ความรู้เป็นฐาน (knowledge-based WSD)

วิธีการนี้จะมีข้อจำกัดเรื่องคำเป้าหมายที่จะใช้ ซึ่งจะต่างจากวิธีการที่ต้องใช้คลังข้อมูล (corpus-based technique) ดังเช่นวิธีการให้เครื่องเรียนรู้แบบมีผู้สอนที่จะศึกษาได้เฉพาะคำที่ได้รับ การกำกับข้อมูลไว้แล้วเท่านั้น (Mihalcea, 2007) ในอดีตวิธีการนี้ใช้กฎทางภาษาที่ผู้เชี่ยวชาญเขียน ขึ้นในการแยกความกำกวมของความหมาย แต่วิธีการนี้ก็มีข้อเสียว่าต้องใช้เวลาในการให้ผู้เชี่ยวชาญ เขียนกฎค่อนข้างมาก อีกทั้งนำไปประยุกต์ใช้กับภาษาอื่นค่อนข้างยาก หรือแม้แต่ในภาษาเดียวกันแต่ เนื้อหาของข้อมูลทีวิเคราะห์เป็นคนละเรื่องกัน กฎที่ใช้แยกความกำกวมก็อาจแตกต่างกันได้ เพราะ คำศัพท์บางคำจะใช้ความหมายแตกต่างกันไปในแต่ละเรื่องหรือหัวข้อ เช่น คำว่า *bench* ใน ภาษาอังกฤษ หากอยู่ในเรื่องทั่ว ๆ ไปก็จะมีความหมายว่าม้านั่งธรรมดา แต่หากเป็นเรื่องที่เกี่ยวข้อง ศาล *bench* จะหมายถึงผู้พิพากษาหรือที่นั่งของผู้พิพากษา เป็นต้น

วิธีการอื่นที่ใช้ความรู้เป็นฐาน เช่น Lesk algorithm ซึ่งเป็นอัลกอริทึมแรกๆ ที่ริเริ่มใช้ ความหมายในพจนานุกรมมาแยกความกำกวม วิธีการนี้มีสมมติฐานว่าคำต่าง ๆ ที่ปรากฏร่วมกันต่าง มีความสัมพันธ์กันและความสัมพันธ์นี้สามารถหาได้จากความหมายของคำนั้น ๆ โดยดูจากความหมาย ที่เหลื่อมหรือทับซ้อน (overlap) กันอยู่ วิธีนี้จึงเป็นการเอาความหมายของคำเป้าหมายไป เปรียบเทียบกับความหมายของคำอื่น ๆ ที่อยู่รอบ ๆ คำเป้าหมาย วิธีนี้สามารถแก้ปัญหาความ กำกวมของคำได้ถูกต้อง 50-70% จากข้อมูลทดสอบขนาดเล็ก (Ide and Véronis, 1998 อ้างถึงใน Edmonds, 2006) แต่วิธีนี้ก็มีปัญหาว่าเหมาะสำหรับใช้หาส่วนที่เหลื่อมกันทางความหมายของคำได้ เพียงไม่กี่คำเท่านั้น เพราะถ้ายังหาส่วนที่เหลื่อมกันของคำจำนวนมาก เวลาที่ต้องใช้ในการวิเคราะห์ก็ จะมากตามไปด้วย (Edmonds, 2006)

อีกตัวอย่างหนึ่ง คือ การหาค่าความคล้ายคลึงกันทางความหมาย (semantic similarity) วิธีนี้มีสมมติฐานว่า คำที่อยู่ในบริบทที่คล้ายกันมักมีความหมายใกล้เคียงกัน ดังนั้นความหมายจึง สามารถหาได้จากการเปรียบเทียบความหมายแล้วดูว่าความหมายใดมีค่าความห่างกันน้อยที่สุดก็จะ เลือกความหมายนั้น (Rada et al., 1989 อ้างถึงใน Mihalcea, 2007) วิธีที่ใช้สำหรับคำนวณหาค่า

ความคล้ายคลึงทางความหมายมีหลายวิธีด้วยกัน เช่น วิธีของ Resnik วิธีของ Lin เป็นต้น อย่างไรก็ตาม วิธีนี้มีข้อด้อยคล้ายกับวิธี Lesk algorithm คือไม่เหมาะจะนำมาใช้กับคู่คำมากกว่า 2 คำขึ้นไป เพราะจะต้องใช้เวลาคำนวณค่อนข้างมาก

วิธีใช้ความรู้เป็นฐานมีข้อดีคือขั้นตอนไม่มีความซับซ้อนมากแต่การจะใช้วิธีนี้ได้ในภาษานั้นจำเป็นต้องมีข้อมูลความรู้หรือพจนานุกรมที่ดีพอเพื่อที่จะนำมาใช้ในการหาความหมายของคำ ซึ่งภาษาส่วนใหญ่ยังขาดแคลนทรัพยากรนี้ก็ยากที่จะใช้วิธีการนี้ได้

3. วิธีการให้เครื่องเรียนรู้แบบไม่มีผู้สอน (Unsupervised WSD)

แนวคิดหลักของวิธีนี้ คือ คำที่มีความหมายคล้ายกันมักปรากฏอยู่ในบริบทที่คล้ายกัน เป็นการให้เครื่องเรียนรู้จากข้อมูลที่ไม่มีการกำกับข้อมูลใด ๆ ไว้ ให้เครื่องแยกความหมายของคำออกมาเองแบบอัตโนมัติ ซึ่งจะต่างจากวิธีแรกและวิธีที่สองที่จะรู้ก่อนว่าคำที่ต้องการศึกษามีความหมายแล้วให้ระบบเลือกความหมายจากความหมายที่มีอยู่แล้ว แต่วิธีนี้จะให้ระบบแยกกลุ่มของข้อความที่คาดว่ามีความหมายใกล้เคียงกันไว้รวมอยู่ด้วยกัน ซึ่งกลุ่มที่แยกจากกันนี้เปรียบเสมือนเป็นความหมายแต่ละความหมาย ดังนั้นจำนวนและความหมายที่ได้จึงค่อนข้างจะแตกต่างจากความหมายที่อยู่ในพจนานุกรม จากที่กล่าวไปจะเห็นวิธีการให้เครื่องเรียนรู้แบบไม่มีผู้สอนมีลักษณะเป็นการแยกความหมายของคำจากข้อมูล จึงเรียกว่า word sense discrimination ในขณะที่ถ้าเป็นงาน word sense disambiguation ทั่วไปจะเป็นการกำกับความหมายให้กับคำ หรือ sense labeling task วิธีการนี้เหมาะสำหรับภาษาที่มีทรัพยากรไม่มากหรือขาดแคลน ทรัพยากรในที่นี้คือ คลังข้อมูลที่มีการกำกับความหมายไว้แล้ว ซึ่งภาษาส่วนใหญ่จะอยู่ในกลุ่มนี้ (Nasiruddin, 2013)

เทคนิควิธีการที่ใช้สามารถแบ่งตามลักษณะการแยกความหมายที่อัลกอริทึมใช้ออกเป็น 3 ประเภท (Navigli, 2012; Nasiruddin, 2013) ดังนี้

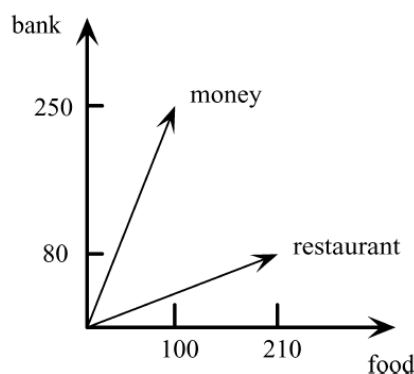
1. วิธีการแยกตามกลุ่มคำ (Word Clustering)

วิธีการนี้จะแยกกลุ่มของคำที่อยู่รอบ ๆ คำเป้าหมาย เช่น คำว่า *plant* มีคำรอบข้าง ได้แก่ *factory, facility, refinery, shrub, perennial, bulb* ซึ่งคำเหล่านี้ต้องมีความหมายใกล้เคียงกับคำเป้าหมายด้วย จากนั้นนำเอาเวกเตอร์คุณสมบัติของคำต่าง ๆ เหล่านี้มาคำนวณ ว่ามีคำใดที่คล้ายกันโดยดูระยะความห่างของเวกเตอร์ คำใดที่คล้ายกันก็จัดเข้าอยู่ในกลุ่มเดียวกัน ซึ่งจากตัวอย่างที่ยกมา เมื่อคำนวณหาค่าความคล้ายคลึงกันทางความหมายแล้วจะสามารถแบ่งออกได้เป็น 2

กลุ่ม คือ (factory, facility, refinery) และ (shrub, perennial, bulb) คำแต่ละกลุ่มแทนแต่ความหมายของ *plant* (Edmonds, 2006) ตัวอย่างงานวิจัยที่ใช้วิธีนี้ เช่น งานของ Lin (1998 อ้างถึงใน Niruddin, 2013) แยกกลุ่มของคำที่คล้ายกับคำเป้าหมาย โดยใช้อัลกอริทึมคำนวณค่าทางสถิติของการพึ่งพากันทางวากยสัมพันธ์ (syntactic dependency statistics) ระหว่างคำ เช่น subject-verb, verb-object, adjective-noun เป็นต้น ระหว่างคำ คำใดที่มีลักษณะการพึ่งพาทางวากยสัมพันธ์เหมือนกันมาก ก็มีแนวโน้มว่าจะเป็นคำที่มีความหมายคล้ายกันมาก

2. วิธีการแยกตามกลุ่มบริบท (Context Clustering)

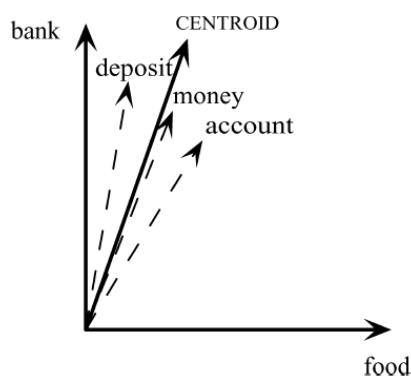
วิธีนี้คล้ายกับวิธีการแยกกลุ่มในวิธีแรก แต่จะดูทั้งบริบท โดยทุกบริบทที่มีคำเป้าหมายอยู่จะแสดงออกมาในรูปของเวกเตอร์บริบท (context vector) แล้วจากนั้นก็นำเวกเตอร์บริบทมาคำนวณหาความคล้ายคลึงกันเพื่อจัดเป็นกลุ่ม ๆ แต่ละกลุ่มหมายถึงแต่ละความหมายของคำเป้าหมาย แนวคิดนี้ได้มาจากเรื่อง word space (Schütze, 1992 อ้างถึงใน Navigli, 2009) คือ เป็น vector space ที่มีคำเป็นขนาดของ space คำสามารถแสดงออกมาในรูปของเวกเตอร์ได้ โดยการใช้ความสัมพันธ์ของการปรากฏร่วมกันของคำ โดยที่คำทั้งสองคำนี้อยู่ภายในขอบเขตของบริบทที่กำหนดไว้ เช่น ประโยคหรือย่อหน้า ภาพต่อไปนี้จะแสดงตัวอย่างของเวกเตอร์คำ



ภาพที่ 2.4 ตัวอย่างของเวกเตอร์คำของคำว่า *restaurant* และ *money* (Navigli, 2009)

จากภาพที่ 2.4 แสดงตัวอย่างของเวกเตอร์คำของคำว่า *restaurant* และ *money* โดยที่เวกเตอร์ของ *restaurant* = (210,80) และ *money* = (100,250) ตัวเลขตัวแรกหมายถึงจำนวนครั้งที่ *restaurant* และ *money* ปรากฏร่วมกับคำว่า *food* และตัวเลขที่สองหมายถึงจำนวนครั้งที่ปรากฏร่วมกับคำว่า *bank* ความคล้ายคลึงกันระหว่างคำสองคำสามารถวัดได้ค่า cosine ระหว่างสองเวกเตอร์ จากนั้นหาค่าเฉลี่ยปกติ (normalized average) ของเวกเตอร์คำต่าง ๆ ที่มีค่า

คล้ายคลึงกันใกล้เคียงกันที่อยู่ในบริบทเดียวกัน เวกเตอร์ค่าเฉลี่ยปกติเรียกว่า centroid ดังแสดงไว้ในภาพที่ 2.5

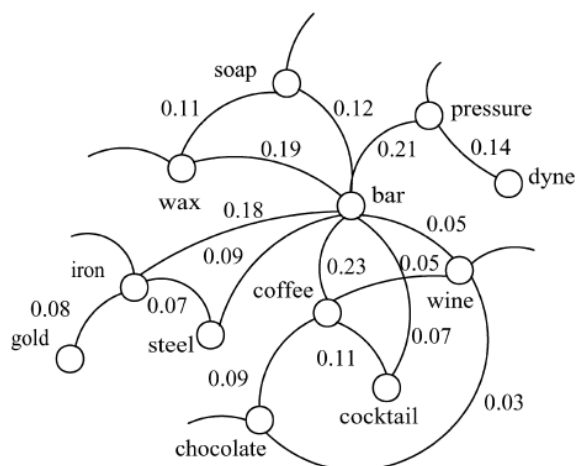


ภาพที่ 2.5 ตัวอย่าง centroid ของคำว่า *stock* ที่ได้จากเวกเตอร์ค่าที่ปรากฏอยู่ในบริบทเดียวกัน (Navigli, 2009)

เมื่อได้ centroid ของบริบทต่าง ๆ แล้ว จะใช้อัลกอริทึมในการช่วยแยกกลุ่มของ centroid อีกครั้งหนึ่ง ตัวอย่างอัลกอริทึม เช่น context-group discrimination ของ Schütze (1998) และค่าเฉลี่ยของกลุ่ม centroid แต่ละกลุ่มแทนความหมายแต่ละความหมายของคำ

3. กราฟคำปรากฏร่วม (Co-occurrence Graph)

วิธีนี้จะคล้ายกับการจับกลุ่มตามคำ แต่จะเปลี่ยนเป็นการสร้างกราฟแสดงการปรากฏร่วมกันของคำที่อยู่ภายในขนาดข้อความที่กำหนด เช่น ย่อหน้า โดยโหนด (nodes/vertices) แต่ละโหนดคือคำที่อยู่ในข้อความ และเส้นเชื่อม (edges) จะแสดงความสัมพันธ์ระหว่างคำต่าง ๆ ซึ่งคำที่จะเลือกมานั้นอาจเลือกมาจากการใช้ความสัมพันธ์ทางไวยากรณ์ระหว่างคำหรือเป็นคำที่ปรากฏร่วมกับคำเป้าหมาย ตัวอย่างกราฟคำปรากฏร่วมแสดงไว้ในภาพที่ 2.6



ภาพที่ 2.6 ส่วนหนึ่งของกราฟคำปรากฏร่วมของคำว่า *bar* (Navigli, 2009)

จากภาพ แต่ละเส้นเชื่อมแสดงค่าน้ำหนักของความสัมพันธ์ในการปรากฏร่วมกันของคำ คำที่มีความถี่ในการปรากฏร่วมกันสูงจะมีค่าเข้าใกล้ 0 แต่ถ้าคำมีความถี่ในการปรากฏร่วมกันต่ำจะมีค่าเข้าใกล้ 1 จากนั้นอัลกอริทึมจะเลือกเอาเฉพาะโหนดที่มีค่าความสัมพันธ์กันสูงขึ้นมาสร้างเป็นเวกเตอร์ และคำนวณผลรวมหาค่าที่ได้สูงสุด ค่าที่ได้สูงสุดแทนความหมายของคำเป้าหมาย

นอกจากเทคนิควิธีที่กล่าวไปแล้วทั้ง 3 วิธี การเรียนรู้แบบไม่มีผู้สอนยังอาจใช้วิธีการแปล (Translation-oriented Approach) ได้ด้วย (Pedersen, 2007) คลังข้อมูลที่ใช้จะเป็น parallel corpus ที่มีการจับคู่คำระหว่างภาษา เพื่อดูว่าคำในภาษาหนึ่งแปลเป็นอีกภาษาหนึ่งว่าอะไร โดยมีแนวคิดที่ว่าคำในภาษาหนึ่งที่มีหลายความหมาย เมื่อแปลเป็นอีกภาษาหนึ่งแล้วความหมายที่ต่างกันจะแสดงถึงความหมายที่ต่างกัน

วิธีการให้เครื่องเรียนรู้แบบไม่มีผู้สอนมีข้อดีคือไม่มีข้อจำกัดเรื่องคลังข้อมูลสามารถนำเอาข้อมูลที่หลากหลายมาเป็นคลังข้อมูลสำหรับฝึกฝนได้ เช่น ข้อมูลจากทวิตเตอร์ (Gella et al., 2013) เป็นต้น รวมถึงเหมาะกับการนำไปใช้กับภาษาที่ขาดแคลนคลังข้อมูลที่มีการกำกับความหมายของคำไว้แล้ว ซึ่งภาษาส่วนใหญ่ในโลกอยู่ในกลุ่มนี้รวมถึงภาษาไทยด้วย แต่เนื่องจากวิธีนี้เป็นวิธีที่ไม่มีผู้สอนให้ระบบก่อน รวมถึงถ้าเป็นภาษาที่ขาดแคลนทรัพยากรทางภาษามาก ๆ ประสิทธิภาพของระบบก็จะยังไม่ได้เท่าที่ควร

2.5 Latent Semantic Analysis (LSA)

Latent Semantic Analysis หรือ Latent Semantic Indexing เป็นวิธีการแสดงความสัมพันธ์คำปรากฏร่วมวิธีหนึ่ง Deerwester และคณะ (Deerwester et al., 1990) เป็นผู้พัฒนา

ขึ้นเพื่อนำมาใช้ในการค้นคืนสารสนเทศ (information retrieval) LSA มีแนวทางหลักว่าให้สกัดเอาคุณสมบัติที่ซ่อน (latent) อยู่ในความหมายของคำออกมา ทั้งนี้โดยมีความเชื่อว่าคำที่มีความหมายคล้ายกันจะปรากฏอยู่ในเอกสารที่มีลักษณะคล้ายกันด้วย นอกจากการหาความคล้ายคลึงกันระหว่างเอกสารแล้ว LSA ยังนำมาใช้หาความคล้ายคลึงกันของคำ ซึ่งการหาความคล้ายคลึงกันของคำนี้เป็นวิธีที่นำมาใช้ในงานวิจัยนี้ เพื่อหาความคล้ายคลึงกันของคำในบริบทต่าง ๆ

LSA จะสร้างเมทริกซ์ของคำปรากฏร่วม โดยคำที่แสดงในแถวและคอลัมน์จะเหมือนกัน ดังนั้นเมทริกซ์ของคำปรากฏร่วมจึงเป็นเมทริกซ์ที่สมมาตร ที่มีจำนวนแถวและคอลัมน์เท่ากัน เช่น

ตารางที่ 2.1 รูปแบบเมทริกซ์ของ LSA (Kontostathis and Pottenger, 2002)

	human	interface	computer	user	system	response	time	EPS	Survey	trees	graph	minors
human	x	1	1	0	2	0	0	1	0	0	0	0
interface	1	x	1	1	1	0	0	1	0	0	0	0
computer	1	1	x	1	1	1	1	0	1	0	0	0
user	0	1	1	x	2	2	2	1	1	0	0	0
system	2	1	1	2	x	1	1	3	1	0	0	0
response	0	0	1	2	1	x	2	0	1	0	0	0
time	0	0	1	2	1	2	x	0	1	0	0	0
EPS	1	1	0	1	3	0	0	x	0	0	0	0
Survey	0	0	1	1	1	1	1	0	x	0	1	1
trees	0	0	0	0	0	0	0	0	0	x	2	1
graph	0	0	0	0	0	0	0	0	1	2	x	2
minors	0	0	0	0	0	0	0	0	1	1	2	x

จากตาราง จะเห็นว่ารายการคำของแถวและคอลัมน์ของเมทริกซ์จะเหมือนกัน โดยคำเหล่านี้จะไม่ใช้คำไวยากรณ์หรือสัญลักษณ์ต่าง ๆ ส่วนตัวเลขที่ปรากฏอยู่ในตารางคือจำนวนครั้งที่คำนั้นปรากฏร่วมกัน ซึ่งเมื่อนับการปรากฏร่วมกันทั้งหมดของคำต่าง ๆ แล้วจะได้เมทริกซ์ขนาดใหญ่ออกมา แล้วค่าความถี่แต่ละค่าจะถูกแปลงเป็นค่าน้ำหนักของคำแทน จากนั้น LSA จะใช้เทคนิคทางคณิตศาสตร์ Singular Value Decomposition (SVD) ในการวิเคราะห์องค์ประกอบของเมทริกซ์ โดยที่จากเดิมที่มีเมทริกซ์เดียวสามารถแยกย่อยออกมาได้เป็น 3 เมทริกซ์ คือ เมทริกซ์ของแถวแนวนอน เมทริกซ์ของคอลัมน์แนวตั้ง และเมทริกซ์แนวทแยง (diagonal matrix) เพื่อเก็บค่าขนาดของเมทริกซ์ไว้ เพื่อที่ว่าหากนำเมทริกซ์ทั้งสามมารวมกันแล้วจะทำให้สามารถวิเคราะห์ย้อนขึ้นไป (reconstruct) หาเมทริกซ์ดั้งเดิมได้ SVD สามารถเขียนแทนได้ด้วยสูตร

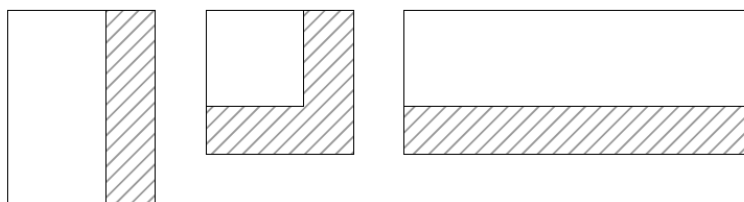
$$M = UDV$$

โดยที่ M คือ เมทริกซ์เดิมที่ยังไม่ผ่านกระบวนการแยกส่วนประกอบ

U และ V คือเมทริกซ์ตั้งฉาก (orthogonal matrix) โดยที่เมทริกซ์หนึ่งแสดงค่าของแถว และอีกเมทริกซ์หนึ่งแสดงค่าของคอลัมน์

D คือ เมทริกซ์แนวทแยงของเมทริกซ์เดิม

ลักษณะการแยกส่วนประกอบของเมทริกซ์ของ SVD แสดงไว้ในภาพที่ 2.7



ภาพที่ 2.7 ลักษณะการแยกองค์ประกอบของเมทริกซ์ด้วย SVD (Clarke, 2007)

จากภาพ จะเห็นว่าข้อมูลของแต่ละเมทริกซ์ถูกแบ่งออกเป็น 2 ส่วน คือส่วนที่เป็นพื้นที่สีขาว กับส่วนที่มีการระบายไว้ เพื่อแสดงให้เห็นว่า SVD จะไม่ใช่ข้อมูลทั้งหมดของเมทริกซ์ที่แยกองค์ประกอบออกมา จากภาพส่วนที่ SVD ใช้คือส่วนที่เป็นพื้นที่สีขาว ดังนั้นขนาดของเมทริกซ์เมื่อผ่านขั้นตอนการแยกองค์ประกอบด้วย SVD แล้วขนาดของเมทริกซ์จะลดลง โดยส่วนที่ลดลงนั้นถือว่าเป็นส่วนที่ไม่มีประโยชน์หรือเป็น noise ของข้อมูล จากนั้นค่าที่ได้จากการแยกองค์ประกอบนี้จะนำมารวมเข้าด้วยกันด้วยการหาค่าผลรวมเชิงเส้น (linear combination) ตัวอย่างของข้อมูลในตารางที่ 2.1 เมื่อผ่านการแยกองค์ประกอบแล้วนำมาหาค่าผลรวมเชิงเส้นจะได้ค่าใหม่ดังตารางที่ 2.2

ตารางที่ 2.2 ค่าใหม่ของค่าจากตารางที่ 1 เมื่อผ่านการแยกองค์ประกอบด้วย SVD (Kontostathis and Pottenger, 2002)

	human	interface	computer	user	system	response	time	EPS	Survey	trees	graph	minors
human	0.62	0.54	0.56	0.94	1.69	0.58	0.58	0.84	0.32	-0.32	-0.34	-0.25
interface	0.54	0.48	0.52	0.87	1.50	0.55	0.55	0.73	0.35	-0.20	-0.19	-0.14
computer	0.56	0.52	0.65	1.09	1.67	0.75	0.75	0.77	0.63	0.15	0.27	0.20
user	0.94	0.87	1.09	1.81	2.79	1.25	1.25	1.28	1.04	0.23	0.42	0.31
system	1.69	1.50	1.67	2.79	4.76	1.81	1.81	2.30	1.20	-0.47	-0.39	-0.28
response	0.58	0.55	0.75	1.25	1.81	0.89	0.89	0.80	0.82	0.38	0.56	0.41
time	0.58	0.55	0.75	1.25	1.81	0.89	0.89	0.80	0.82	0.38	0.56	0.41
EPS	0.84	0.73	0.77	1.28	2.30	0.80	0.80	1.13	0.46	-0.41	-0.43	-0.31
Survey	0.32	0.35	0.63	1.04	1.20	0.82	0.82	0.46	0.96	0.88	1.17	0.85
trees	-0.32	-0.20	0.15	0.23	-0.47	0.38	0.38	-0.41	0.88	1.55	1.96	1.43
graph	-0.34	-0.19	0.27	0.42	-0.39	0.56	0.56	-0.43	1.17	1.96	2.50	1.81
minors	-0.25	-0.14	0.20	0.31	-0.28	0.41	0.41	-0.31	0.85	1.43	1.81	1.32

หากเปรียบเทียบตารางที่ 2.1 และตารางที่ 2.2 จะเห็นว่าในตารางที่ 2.2 มีการกำหนดค่าให้กับคำทุกคำ ซึ่งจะต่างจากตารางที่ 2.1 ที่คำปรากฏร่วมบางคู่มีค่าเป็น 0 เนื่องจากค่าทั้งสองค่าไม่ได้ปรากฏร่วมกัน ทั้งนี้เพราะ SVD จะไปคำนวณค่าจากการปรากฏร่วมกับคำอื่น ๆ แล้วจึงมากำหนดค่าใหม่ให้คำนั้น ๆ ว่ามีความสัมพันธ์กับคำอื่นอย่างไร ซึ่งลักษณะเช่นนี้คือการแสดงความสัมพันธ์ที่ซ่อน (latent) อยู่ออกมา ตัวอย่างเช่น คำว่า *human* กับ *user* ไม่เคยปรากฏร่วมกัน ดังจะเห็นว่า *human* กับ *user* มีความถี่เป็น 0 ในตารางที่ 2.1 แต่เนื่องจากทั้งสองค่าต่างปรากฏร่วมกับ *interface* เหมือนกัน ดังนั้นความสัมพันธ์ของ *human* กับ *user* จึงมีลักษณะเป็นการปรากฏร่วมกันในลำดับที่ 2 (second order co-occurrence) คือไม่ได้ปรากฏร่วมกันโดยตรง แต่ต่างไปปรากฏร่วมกับคำอื่นซึ่งเป็นคำเดียวกัน ดังนั้นในตารางที่ 2.2 *human* กับ *user* จึงมีค่าความสัมพันธ์เท่ากับ 0.94 เป็นต้น

มีงานวิจัยที่ผ่านมาหลายงานที่แสดงให้เห็นว่า LSA มีประสิทธิภาพในการหาความหมายของคำ เช่น Dumais (1991) ทดลองโดยใช้ LSI ในการค้นคืนสารสนเทศ (information retrieval) พบว่า LSI สามารถจับคู่เอกสารได้ตรงกับความหมายที่ผู้ใช้ (user) ต้องการแม้ว่าในเอกสารจะใช้คำที่แตกต่างออกไปจากคำที่ใช้ค้นหา หรือการทดสอบเปรียบเทียบประสิทธิภาพของ LSA กับมนุษย์ในการหาคำพ้องความหมาย (synonym) โดยคำที่ใช้ทดสอบมาจากแบบทดสอบที่ไม่ใช้แล้วของ Educational Testing Service (ETS) และ Test of English as a Foreign Language (TOEFL) ผลที่ออกมาพบว่า LSA ตอบได้ถูกต้อง 65% ซึ่งตรงกับคะแนนเฉลี่ยของนักเรียนส่วนใหญ่ที่เข้าสอบเพื่อศึกษาต่อในระดับอุดมศึกษาในประเทศสหรัฐอเมริกาและมาจากประเทศที่ไม่ได้ใช้ภาษาอังกฤษเป็นภาษาแม่ (Landauer et al., 1998) งานที่เกี่ยวกับการแยกความหมายของคำ เช่น Levin et al. (2006) ใช้ LSA ในการแยกความหมายของคำโดยให้เครื่องเรียนรู้แบบไม่มีผู้สอน โดยให้ LSA คำนวณเวกเตอร์บริบทที่อยู่รอบ ๆ คำที่มีความหมายกำกวม คำที่ศึกษาได้แก่ *hard* (คำคุณศัพท์), *interest* (คำนาม), *line* (คำนาม), *serve* (คำกริยา) แล้วจากนั้นให้เครื่องจัดกลุ่มบริบทแยกตามความหมายของคำโดยดูค่าความห่างระหว่างเวกเตอร์ ถ้าค่าใกล้เคียงกันก็จัดให้อยู่กลุ่มเดียวกัน มีความหมายใกล้เคียงกัน แต่ถ้าค่าห่างกันก็จัดให้เป็นคนละกลุ่ม ซึ่งกลุ่มต่าง ๆ ที่แยกกันนี้แทนความหมายแต่ละความหมาย ผลที่ออกมาพบว่า LSA สามารถแยกกลุ่มความหมายของคำทุกคำได้ถูกต้องสูงกว่าค่าพื้นฐาน (baseline) ที่ตั้งไว้ นอกจากนี้ยังมีงานของ Pino และ Eskenazi (2009) ใช้ LSA ในการแยกความหมายของคำในการเรียนคำศัพท์ภาษาอังกฤษ โดยแบบทดสอบเป็นลักษณะการเลือกคำจากตัวเลือกที่ให้มาเพื่อนำไปเติมลงในช่องว่างให้ถูกต้อง โดยในการทดลองนี้ได้เปรียบเทียบกับวิธี Lesk algorithm ด้วย ผลปรากฏว่า LSA สามารถเลือกคำตอบออกมาได้ถูกต้องมากกว่า Lesk algorithm

2.6 การศึกษาความกำกวมทางความหมายในภาษาไทย

งานวิจัยที่ศึกษาเกี่ยวกับการแก้ปัญหาความกำกวมทางความหมายในภาษาไทยที่ผ่านมายังมีไม่มากนัก โดยมีงานของ Kanokrattananukul (2001) ใช้ Decision List ในการแยกความกำกวมทางความหมายของคำนาม หัว และคำกริยา เก็บ คุณสมบัติที่ใช้ช่วยในการแยกความหมายคือคำบริบทที่ปรากฏร่วมกับคำเป้าหมาย ซึ่งในงานวิจัยพบว่า คำบริบท 5 ตำแหน่งจากคำเป้าหมายเพียงพอในการแยกความกำกวมของคำ และพบว่าบริบทที่ใช้ช่วยแยกความหมายของทั้ง 2 คำ ส่วนใหญ่จะเป็นบริบททางด้านขวา อย่างไรก็ตามงานของ Kanokrattananukul เป็นลักษณะให้เครื่องเรียนรู้จึงต้องมีการสร้างคลังข้อมูลสำหรับฝึกฝนและทดสอบ หากคลังข้อมูลสำหรับฝึกฝนมีมากขึ้นก็จะช่วยให้ประสิทธิภาพของระบบดีขึ้นตามไปด้วย แม้ผลการทดสอบประสิทธิภาพของระบบจะดีแต่การสร้างคลังข้อมูลมีข้อด้อยว่าต้องใช้เวลาค่อนข้างมาก

Pongpinigpinyo and Rivepiboon (2005) ได้ใช้วิธี Latent Semantic Indexing (LSI) เป็นการให้เครื่องเรียนรู้แบบไม่มีผู้สอน งานวิจัยนี้ศึกษาเปรียบเทียบประสิทธิภาพของระบบทั้งภาษาไทยและภาษาอังกฤษ โดยคลังข้อมูลภาษาไทยเป็นคลังข้อมูลจากงานวิจัยของ Kanokrattananukul สำหรับภาษาอังกฤษผู้วิจัยใช้คลังข้อมูลของ Computing Research Laboratory ทั้งนี้เพื่อแสดงให้เห็นว่าวิธีการของผู้วิจัยสามารถใช้ได้กับทั้งภาษาไทยและภาษาอังกฤษ คุณสมบัติหลักที่ใช้ในงานวิจัยนี้คือคำปรากฏร่วม ในงานวิจัยนี้ใช้ Singular Value Decomposition (SVD) ในการลดขนาดของเมตริกซ์ของคำปรากฏร่วม เพื่อให้เห็นความสัมพันธ์ของคำปรากฏร่วมต่าง ๆ ได้ชัดเจนขึ้น โดยส่วนที่ถูกตัดออกไปนั้นถือว่าเป็นส่วนที่ไม่สำคัญหรือ noise ของข้อมูล ซึ่งจากผลการวิจัยพบว่าค่าความถูกต้อง (F-measure) ของระบบอยู่ที่ประมาณ 70%

นอกจากนี้ ยังมีงานวิจัยของ Modhiran และคณะ (Modhiran et al., 2004) ที่ศึกษาเปรียบเทียบวิธีการเรียนรู้ด้วยเครื่อง 3 วิธี ได้แก่ SNOW, Naïve Bayes และ Support Vector Machine (SVM) และใช้คำในการศึกษาทั้งหมด 10 คำ ในงานวิจัยนี้ได้ใช้หน้าที่ของคำเป็นคุณสมบัติหนึ่งในการช่วยแยกความหมายด้วย ซึ่งจากผลการวิจัยพบว่าค่าความถูกต้องของคำแต่ละคำ และของแต่ละระบบไม่เท่ากัน นั้นแสดงให้เห็นว่าประสิทธิภาพของระบบขึ้นอยู่กับการเลือกมาศึกษาด้วย โดยในงานวิจัยนี้ SVM มีประสิทธิภาพมากที่สุดในการช่วยแก้ความกำกวมของคำ และคำที่ระบบสามารถแยกความหมายได้ถูกต้องมากที่สุดคือ กิน

จากการทบทวนวรรณกรรมต่าง ๆ ทำให้ผู้วิจัยเข้าใจธรรมชาติของคำหลายความหมายได้ดีขึ้น รวมถึงเห็นแนวทางในการวิเคราะห์ความหมายของคำหลายนัยความหมาย ว่าเมื่อใดจึงควรแยก

ความหมายออกเป็นความหมายใหม่ หรือยังไม่ควรแยก นอกจากนี้ ยังทำให้เข้าใจหลักการและขั้นตอนการประมวลผลของ LSA ซึ่งเป็นวิธีการที่จะนำมาใช้งานวิจัยนี้ได้ดียิ่งขึ้น

บทที่ 3

วิธีดำเนินการวิจัย

บทนี้ผู้วิจัยได้แบ่งเนื้อหาออกเป็น 4 ส่วน ได้แก่ วิธีการเลือกคำที่ใช้ในงานวิจัย วิธีการรวบรวมข้อมูล วิธีการทดลอง และส่วนสุดท้ายคือวิธีการประเมินประสิทธิภาพของระบบ

3.1 วิธีการเลือกคำที่ใช้ในงานวิจัย

ในงานวิจัยนี้ ผู้วิจัยเลือกคำที่ใช้เป็นตัวแทนของคำนามสองคำ และคำที่เป็นตัวแทนของคำกริยาสองคำ โดยที่คำหนึ่งมีนัยความหมายมากในขณะที่อีกคำมีนัยความหมายน้อย คำนาม 2 คำที่ผู้วิจัยใช้ในงานนี้คือ *เสียง* และ *หัว* และคำกริยา 2 คำ คือ *บอก* และ *ติด* โดยคู่คำที่เป็นตัวแทนของคำที่มีนัยความหมายน้อย ได้แก่ *เสียง* (นาม) และ *บอก* (กริยา) และคู่คำที่เป็นตัวแทนของคำที่มีนัยความหมายมาก ได้แก่ *หัว* (นาม) และ *ติด* (กริยา) คำทั้ง 4 คำนี้ คัดเลือกมาจากรายการคำ 5,000 คำแรกที่ปรากฏอยู่ในคลังข้อมูลภาษาไทยแห่งชาติ (Thai National Corpus: TNC) และคัดเลือกเฉพาะคำที่มีหน้าที่ของคำ (part of speech) เพียงหน้าที่เดียว เนื่องจากความถี่ที่แสดงในรายการคำ 5,000 คำนี้เป็นการรวบรวมตามการปรากฏของรูปคำเท่านั้น โดยไม่ได้นำหน้าที่ของคำมาร่วมพิจารณาด้วย ดังนั้นหากคำที่เลือกมามีหลายหน้าที่ ปริมาณข้อมูลที่เห็นก็จะต้องแบ่งกระจายไปตามแต่ละหน้าที่ของคำ ซึ่งการจะรู้จำนวนตัวอย่างที่แน่นอนก็ต้องใช้เวลาในการคัดแยกคำแต่ละหน้าที่ด้วย ดังนั้นเพื่อความสะดวกในการหาคำตอบที่ต้องการ ผู้วิจัยจึงเลือกคำที่มีหน้าที่เดียว เพื่อให้รู้ปริมาณตัวอย่างแน่นอน และเป็นคำที่อยู่ในลำดับต้น ๆ ของรายการคำ เนื่องจากรายการนี้เรียงลำดับตามความถี่ที่คำ ๆ นั้นปรากฏในข้อมูล นั้นหมายความว่าหากคำใดมีความถี่สูง ปริมาณตัวอย่างของคำ ๆ นั้นก็มากตามไปด้วย เพื่อให้ได้ปริมาณตัวอย่างที่เพียงพอที่จะนำมาใช้ในการประมวลผลข้อมูล ผู้วิจัยจึงเลือกคำในลำดับต้น ๆ เพราะไม่ใช่ข้อมูลทั้งหมดจะใช้ได้เสมอไป ซึ่งรายละเอียดวิธีการรวบรวมข้อมูลนี้จะกล่าวถึงในหัวข้อถัดไป นอกจากปัจจัยเรื่องความถี่แล้วอีกปัจจัยหนึ่งที่ผู้วิจัยนำมาใช้คัดเลือกคำคือ ปริมาณความหมายของคำ เนื่องจากงานวิจัยนี้ต้องการศึกษาเปรียบเทียบประสิทธิภาพของระบบระหว่างคำที่มีนัยความหมายน้อยกับคำที่มีนัยความหมายมาก โดยที่คำที่มีนัยความหมายน้อย ผู้วิจัยจะคัดเลือกคำที่มีนัยความหมายน้อยกว่า 5 ความหมาย และคำที่มีนัยความหมายมาก คือ คำที่มีนัยความหมายมากกว่า 7 ความหมาย โดยอิงจากจำนวนความหมายในพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554 (ราชบัณฑิตยสถาน, 2556) รวมถึงปริมาณความหมายของคำนามและคำกริยาก็ต้องใกล้เคียงกัน เนื่องจากมีการเปรียบเทียบประสิทธิภาพของระบบระหว่างคำนามและคำกริยา ดังนั้นเพื่อคุมปัจจัยเรื่องปริมาณความหมายที่อาจมีผลต่อ

ประสิทธิภาพของระบบ ผู้วิจัยจึงเลือกคำนามและคำกริยาที่มีปริมาณความหมายใกล้เคียงกันมากที่สุด

ตารางที่ 3.1 และ 3.2 แสดงการเปรียบเทียบปริมาณความหมายของ *เสียง* กับ *บอก* และ *หัว* กับ *ติด* ตามลำดับ ตามที่ปรากฏในพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554 (ราชบัณฑิตยสถาน , 2556) ดังนี้

ตารางที่ 3.1 ความหมายของ *เสียง* และ *บอก* ตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554

เสียง (คำนาม)	บอก (คำกริยา)
1. สิ่งที่ได้รับรู้ได้ด้วยหู เช่น เสียงฟ้าร้อง เสียงเพลง เสียงพูด	1. พูดให้รู้, เล่าให้ฟัง เช่น บอกกันต่อ ๆ มา, บอกอาการให้หมอฟัง
2. ความเห็น เช่น เรื่องนี้ฉันไม่ออกเสียง	2. บ่งให้รู้ เช่น สัญญาณบอกเหตุร้าย อีกาบอกข่าว
3. ความนิยม, คະแนนเสียง เช่น ลงมติกันแล้ว เขาชนะเพราะได้เสียงข้างมาก	3. แนะนำ, สอน เช่น บอกให้ทำอย่างนั้นอย่างนี้, อย่าบอกหนังสือสังฆราช
	4. รายงาน เช่น ไบบอก

ตารางที่ 3.2 ความหมายของ *หัว* และ *ติด* ตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554

หัว 1 (คำนาม)	ติด (คำกริยา)
1. ส่วนบนสุดของร่างกายของคนหรือสัตว์	1. อาการที่ข้องอยู่ เช่น ติดอยู่ในโลก ติดอยู่ในกิเลส
2. ส่วนของพืชพันธุ์บางอย่างตอนที่อยู่ใต้ดิน เช่น หัวหอม หัวผักกาด, ส่วนที่อยู่ใต้ดินของพืชบางชนิด เป็นที่เกิดต้นอ่อน	2. ประดับ เช่น ติดเหรียญ ติดตรา เอาดอกไม้ติดผม
3. ส่วนแห่งสิ่งของบางอย่างที่อยู่ข้างหน้า หรือข้างต้น หรือแรกเริ่ม เรียกว่า หัวของสิ่งนั้น ๆ เช่น หัวเรือ หัวถนน	3. แนบอยู่ เช่น เนื้อติดมัน เนื้อติดกระดูก
4. ช่วงแรกเริ่มของเวลา เช่น หัวปี หัววัน หัวค่ำ	4. ทำให้เกิดขัดอยู่หรือข้องอยู่จนไม่อาจให้เคลื่อนไปได้ เช่น ติดขบวนแห่ ติดไฟแดง ติดฝน
5. ส่วนแห่งสิ่งของที่เป็นยอด เช่น หัวผี, ส่วนแห่งสิ่งของที่ยื่นเด่นออกไป เช่น หัวแหลม หัวสะพาน	5. ขอบอย่างขาดไม่ได้ เช่น ติดบุหรื ติดผืน
	6. อาศัยไปด้วยหรือมาด้วย เช่น ติดรถไปด้วย
	7. ผนึก เช่น ติดแสตมป์ ติดประกาศ ติดกาว

หัว 1 (คำนาม)	ติด (คำกริยา)
6. ในการเล่นปั่นแปะหรือโยนหัวโยนก้อย เรียกสมมุติด้านหนึ่งของเงินปลีกว่า ด้านหัว คู่กับ ด้านก้อย 7. ส่วนที่ตรงข้ามกับหางหรือท้าย เช่น หัวแถว หางแถว หัวเรือ หรือส่วนที่ตรงข้ามกับ กัน ในความว่า หัวหวานกันเปรี้ยว 8. ส่วนที่เป็นแก่นสาร เช่น หัวยา หัวเหล้า 9. ส่วนเริ่มต้นที่เป็นวงของตัวหนังสือ	8. จุด เช่น ติดไฟ ติดตะเกียง 9. ต่อเนื่องอย่างใกล้ชิด เช่น ห้องติดกัน 10. ได้รับเชื้อโรค เช่น ติดหิด 11. คงอยู่หรือทำให้คงอยู่ เช่น สีสติดเสื้อ เสื้อติด สี 12. อยู่ในที่คุมขัง เช่น ติดคุก ติดตะราง 13. (ปาก) มีหนี้สินติดค้างกันอยู่ เช่น ติดเงิน 14. ไม่ร่วง, ไม่หลุด, (ใช้แก่ผลไม้) เช่น มะม่วงปี นี้ติดมาก 15. คล้าย, ใกล้ข้าง เช่น หน้าตาติดไปทางพ่อ
หัว 2 (คำนาม)	
1. สติปัญญา, ความสามารถพิเศษ, ความคิด ริเริ่ม เช่น เด็กคนนี้มีหัวทางดนตรี 2. ผู้ที่มีความคิดหนักไปทางใดทางหนึ่ง เช่น หัวกฎหมาย 3. ปัญญา, ความคิด เช่น หัวดี, หัวไว	

สำหรับคำว่า หัว ในพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 (ราชบัณฑิตยสถาน, 2556) พบว่าคำนาม หัว มี 2 คำด้วยกัน เป็นคำพ้องรูป แต่ผู้วิจัยจะรวมความหมายของทั้ง 2 คำไว้ด้วยกัน เนื่องจากในงานวิจัยนี้ เป็นการศึกษาเรื่องการแยกความหมายเท่านั้น คือสนใจว่ารูปคำหนึ่งมีความหมายอะไรได้บ้างเพื่อนำไปใช้ในการประมวลผลภาษาต่อไป สำหรับคอมพิวเตอร์แล้ว เมื่อพบรูปคำหนึ่ง สิ่งที่ต้องสามารถทำได้คือวิเคราะห์ได้ว่ารูปคำในบริบทนั้น ๆ มีความหมายอะไร โดยไม่สนใจว่าเป็นความหมายที่มาจากคำพ้องรูปหรือมาจากคำหลายความหมาย ดังนั้นเมื่อรวมความหมายของ หัว 1 และ หัว 2 เข้าด้วยกันแล้ว หัวตามพจนานุกรม จึงมี 12 ความหมาย อย่างไรก็ตามความหมายของคำที่ใช้จริงอาจมีเพิ่มเติมแตกต่างจากความหมายที่ระบุไว้ในพจนานุกรมและเราไม่ทราบมาก่อนว่าความหมายทั้งหมดของคำนั้น ๆ มีได้กี่ความหมาย ดังนั้น ผู้วิจัยจึงวิเคราะห์ความหมายของคำเพิ่มเติมจากคลังข้อมูลที่ใช้ ซึ่งจะกล่าวถึงในบทถัดไป

3.2 วิธีการรวบรวมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้มาจากคลังข้อมูลภาษาไทยแห่งชาติ โดยการคัดเลือกข้อมูลแบบสุ่มเพื่อไม่ให้ข้อมูลหนักไปทางข้อมูลชนิดใดชนิดหนึ่งโดยเฉพาะ จากนั้นนำข้อมูลมาผ่านโปรแกรมตัดคำ Thai Word Segmentation 2.1 (วิโรจน์ อรุณมานะกุล, 2545) เนื่องจากงานวิจัยนี้จะใช้บริบทเป็น

ตัวช่วยในการแยกความหมายของคำ ดังนั้นข้อมูลที่ไม่มีบริบททางด้านซ้ายหรือด้านขวา หรือมีคำปรากฏร่วมทางด้านซ้ายหรือขวาน้อยกว่า 7 คำจะไม่นำมาใช้วิเคราะห์ด้วย เช่น

ไป|ลอง|สินค้า|ใน|ห้อง|ลอง|ทั้ง|ที่|มี|ป้าย| |‘|งด|ใช้|ห้อง|ลอง|ชั่วคราว|’| |**ติด**|อยู่| |-----

-----|อาร์|คลา|**หัว**|ป้อย|ๆ| |ก่อน|บน|อย่าง|น่า|สงสาร| |(จะ|น่า|สงสาร|มาก|

นอกจากนี้ กรณีที่คำที่ศึกษาไปเป็นส่วนหนึ่งของคำอื่น เช่น เป็นส่วนหนึ่งของชื่อเฉพาะ หรือเป็นคำที่มีความหมายเป็นสำนวน ก็จะไม่นำมาใช้ในการวิเคราะห์เช่นกัน เช่น

ดิฉัน|เลย|ตัดสินใจ|โทร|หา|ช่าง|ที่|ศูนย์|ฮอน|ด้า| |**หัว**|หมาก| |และ|อธิบาย|ตำแหน่ง|

โทร|ผ่าน|สาย|เคเบิล|ฟรี|ทั่วโลก| |เท่า|นี้|สาย|เคเบิล|ก็|เหมือน|เสื่อ|**ติด**|ปีก|แล้ว|คะ| |-----

จากตัวอย่างแรก หัว เป็นส่วนหนึ่งของชื่อแขวง หัวหมาก ซึ่งเป็นชื่อเฉพาะ และ ติด ในตัวอย่างที่สอง เป็นส่วนหนึ่งของสำนวน เสื่อติดปีก ที่เปรียบกับสถานการณ์ที่ส่งผลให้ผู้ที่กำลังรุ่งเรืองอยู่แล้ว เจริญรุ่งเรืองมากขึ้นไปอีก ในที่นี้คือโปรแกรมสายเคเบิลที่จากเดิมมีผู้นิยมใช้มากอยู่แล้ว ก็จะมีคนมาใช้งานมากขึ้น เนื่องจากสามารถใช้โทรศัพท์ได้ฟรีทั่วโลก เป็นต้น

จากนั้นผู้วิจัยลบข้อมูลส่วนที่เป็นตัวเลขและเครื่องหมายต่าง ๆ เช่น “ ‘ ? ! , . () เป็นต้น ออก รวมถึงตัดข้อมูลส่วนที่เป็นช่องว่างออกด้วยเช่นกัน เช่น

--|"**ติด**|พี|อยู่|พัน|หนึ่ง|.|. |ไว้|เดือน|หน้า|นะ|"| |-----

ถ่ายทำ|ปัญญา| |เรณู| |ภาค| |2| |เขา|**บอก**|ว่า| |"จบ|การ|ถ่ายทำ|ประมาณ| |28| |

เนื่องจากตัวเลขและเครื่องหมายต่าง ๆ ส่วนใหญ่ไม่มีความสัมพันธ์กับคำ กล่าวคือ ไม่ได้เป็นคำปรากฏร่วมกับคำใด ๆ ดังนั้นตัวเลขและเครื่องหมายจึงมีส่วนช่วยในการแยกความหมายของคำได้น้อยมาก จากตัวอย่างจะเห็นว่าเมื่อข้อมูลผ่านโปรแกรมตัดคำแล้ว ตัวเลขและเครื่องหมายจะถูกตัดแยกเปรียบเหมือนเป็นคำ 1 คำ ดังนั้น หากไม่ตัดข้อมูลส่วนนี้ออก ก็จะเกิดเป็นคำปรากฏร่วมในลักษณะ คำ-เครื่องหมาย หรือ คำ-ตัวเลข ได้ สำหรับ ช่องว่าง นั้นแม้จะมีประโยชน์ในการช่วยบ่งชี้จุดสิ้นสุดของข้อความหรือประโยค แต่จากข้อมูลพบว่า ช่องว่าง มีความถี่ในการปรากฏสูงมาก เมื่อได้เรียงลำดับตามความถี่ในการปรากฏร่วมกับคำเป้าหมายแล้ว พบว่าช่องว่างมีความถี่อยู่ในช่วงอันดับ 1-5 ของทุกคำทั้งบริบททางด้านซ้ายและด้านขวา ดังจะเห็นได้จากตารางที่ 3.3-3.6 ที่จะแสดงลำดับของคำปรากฏร่วมเรียงตามความถี่ในบริบททางด้านซ้าย 3 ตำแหน่งและด้านขวา 3 ตำแหน่ง โดยส่วนที่แรเงาหมายถึงช่องว่าง ดังนี้

ตารางที่ 3.3 ความถี่ของคำบริบททางซ้ายและขวา 3 ตำแหน่ง 7 คำแรกของ *เสียง*

คำบริบททางซ้าย						เสียง	คำบริบททางขวา					
ตำแหน่งที่ 3		ตำแหน่งที่ 2		ตำแหน่งที่ 1			ตำแหน่งที่ 1		ตำแหน่งที่ 2		ตำแหน่งที่ 3	
	290		266		382			143		262		252
เสียง	52	การ	44	ได้อื่น	157		ตั้ง	108	ของ	75	ตั้ง	56
จะ	35	จะ	39	เป็น	86		ที่	85	ที่	52	มา	36
ไม่	32	ที่	34	และ	55		ของ	80	ตั้ง	41	ของ	31
ใน	31	ก็	32	ด้วย	55		หัวเราะ	47	ๆ	41	ๆ	27
ที่	30	มา	30	ของ	53		ข้างมาก	32	ใน	37	ไป	26
การ	28	เขา	29	ทำ	49		ร้อง	30	จาก	30	เสียง	24

ตารางที่ 3.4 ความถี่ของคำบริบททางซ้ายและขวา 3 ตำแหน่ง 7 คำแรกของ *บอก*

คำบริบททางซ้าย						บอก	คำบริบททางขวา					
ตำแหน่งที่ 3		ตำแหน่งที่ 2		ตำแหน่งที่ 1			ตำแหน่งที่ 1		ตำแหน่งที่ 2		ตำแหน่งที่ 3	
	382		442		172		ว่า	738		280		203
ที่	49	ที่	67	จะ	149		ให้	118	ว่า	186	ว่า	134
ก็	34	ไม่	60	ก็	129			77	ไม่	58	ไม่	62
แต่	27	ก็	51	เขา	116		กับ	66	เขา	43	จะ	38
ไม่	26	แต่	34	มา	75		ได้	61	เป็น	37	เป็น	36
และ	24	เขา	30	ที่	51		ไม่	59	ให้	36	ให้	33
แล้ว	21	เป็น	27	ผม	40		เขา	44	ถูก	34	ที่	32

ตารางที่ 3.5 ความถี่ของคำบริบททางซ้ายและขวา 3 ตำแหน่ง 7 คำแรกของ *หัว*

คำบริบททางซ้าย						คู่	คำบริบททางขวา					
ตำแหน่งที่ 3		ตำแหน่งที่ 2		ตำแหน่งที่ 1			ตำแหน่งที่ 1		ตำแหน่งที่ 2		ตำแหน่งที่ 3	
	226		202		184			190		326		292
มี	35	มา	66	ใน	76		เข้า	70	และ	36	ๆ	38
ที่	30	ไป	48	ที่	68		ของ	68	ๆ	36	ที่	35
การ	25	อยู่	45	ต่อ	67		ฉัน	48	ไป	32	มี	31
หัว	24	จะ	44	บน	58		เตียง	44	ที่	31	ไม่	30
ก็	23	ที่	35	จาก	38		ฉีดยา	39	เท้า	30	ไป	28
จะ	23	ก็	28	มี	35		และ	34	แล้ว	27	เป็น	28

ตารางที่ 3.6 ความถี่ของคำบริบททางซ้ายและขวา 3 ตำแหน่ง 7 คำแรกของ ติด

คำบริบททางซ้าย						คำ	คำบริบททางขวา					
ตำแหน่งที่ 3		ตำแหน่งที่ 2		ตำแหน่งที่ 1			ตำแหน่งที่ 1		ตำแหน่งที่ 2		ตำแหน่งที่ 3	
	245		191	ที่	150		เชื้อ	273		362		249
ที่	47	มี	65		116		อยู่	219	กับ	84	ที่	41
มี	42	ที่	63	ยึด	103		กับ	198	ใน	70	ใน	38
ฟื้นฟู	38	ไม่	57	การ	99		กัน	98	ที่	66	และ	29
จะ	33	สมรรถภาพ	36	ผู้	89			70	อยู่	43	เป็น	28
และ	30	เป็น	25	ไม่	67		ๆ	65	เป็น	28	กัน	25
เป็น	29	และ	25	มา	67		ยาเสพติด	53	ของ	24	ไม่	25

จากตารางที่ 3.3-3.6 จะเห็นว่า ช่องว่างมีความถี่สูงมาก ซึ่งในงานวิจัยนี้สิ่งสำคัญที่ผู้วิจัยจะนำมาใช้ในการจัดกลุ่มความหมายคือคำที่ปรากฏร่วมกับคำเป้าหมาย ดังนั้นเพื่อให้ได้คำปรากฏร่วมที่เป็นคำจริง ๆ ไม่ใช่เครื่องหมายวรรคตอน ผู้วิจัยจึงจะตัดส่วนที่เป็นช่องว่างทั้งหมดออกจากข้อมูล เพื่อไม่ให้เกิดคู่คำปรากฏร่วมที่เป็น คำ-ช่องว่าง

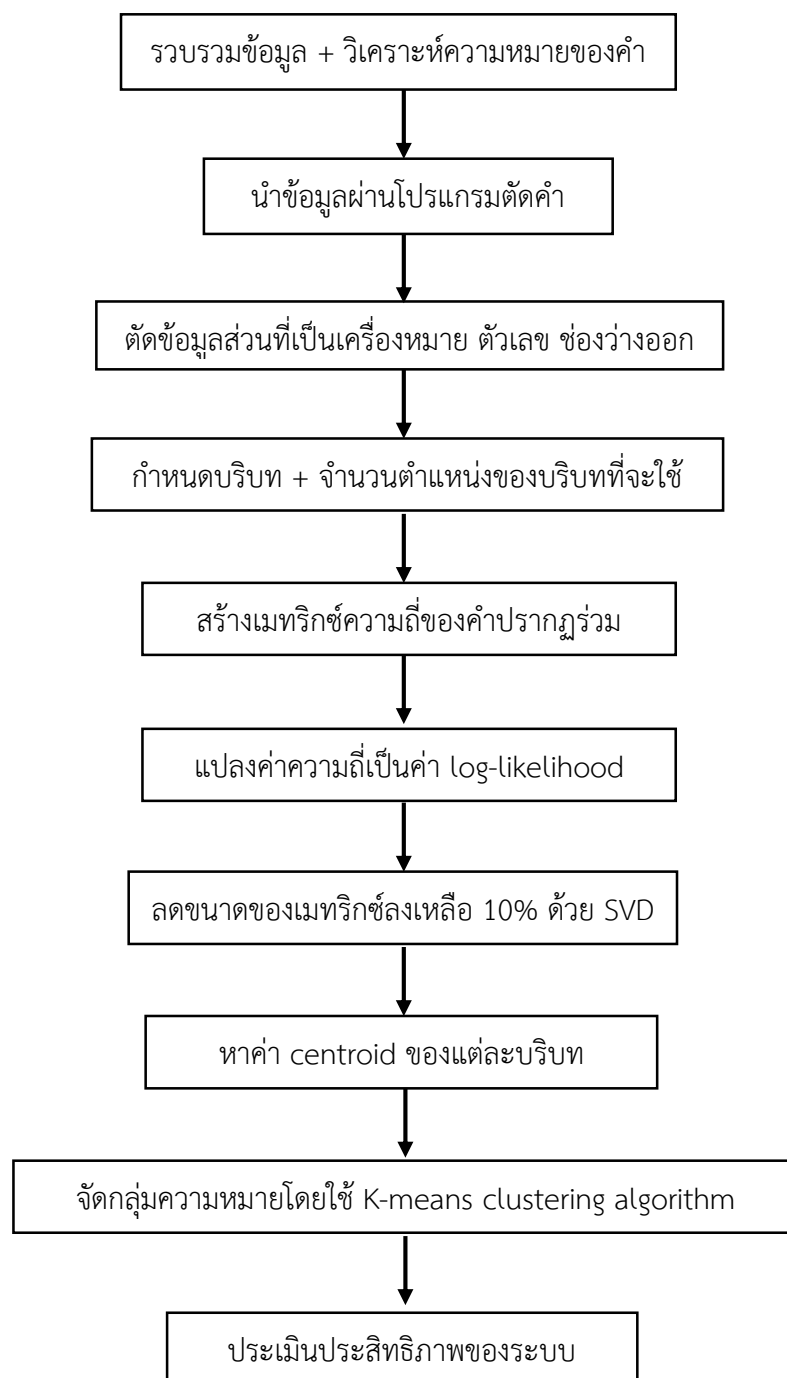
สรุปจำนวนตัวอย่างข้อมูลที่ใช้ของคำที่ศึกษาทั้ง 4 คำ ได้แสดงไว้ในตารางที่ 3.7 ดังนี้

ตารางที่ 3.7 จำนวนตัวอย่างข้อมูลของ เสียง ห้ว บอก ติด

คำนาม		คำกริยา	
เสียง	ห้ว	บอก	ติด
2,018	2,006	2,018	2,006

3.3 วิธีการทดลอง

ขั้นตอนการดำเนินการวิจัยทั้งหมดได้แสดงไว้ในแผนภาพที่ 3.1 ดังนี้



ภาพที่ 3.1 ขั้นตอนของระบบการแยกความหมายของคำ

วิธีการที่ผู้วิจัยจะใช้ในการประมวลผลในงานวิจัยนี้ คือ วิธีการวิเคราะห์ความหมายแอบแฝง (Latent Semantic Analysis : LSA) ของ Deerwester และคณะ (1990) ซึ่ง LSA มีแนวคิดหลักว่า คำที่ปรากฏอยู่ในข้อมูลหรือในบริบทต่าง ๆ นั้นมีคุณสมบัติที่ซ่อน (latent) อยู่ และคุณสมบัติที่ซ่อนอยู่นี้จะช่วยให้สามารถเข้าใจความหมายของคำได้ โดยมีความเชื่อว่าคำที่มีความหมายคล้ายกันมักจะปรากฏอยู่ในบริบทที่มีลักษณะคล้ายกัน ดังนั้นความหมายของคำ ๆ หนึ่งจึงสามารถตีความได้จากบริบท จากแนวคิดของ LSA ดังกล่าว ผู้วิจัยจึงจะดึงคำบริบทที่ปรากฏร่วมกับคำเป้าหมายมาประมวลผล โดยในการดึงคำบริบทนี้ จะดึงเอาคำบริบททางด้านซ้ายอย่างเดียว คำบริบททางด้านขวาอย่างเดียว และคำบริบททั้งด้านซ้ายและด้านขวาที่อยู่ติดกับคำเป้าหมาย โดยจะกำหนดกรอบหน้าต่างหรือระยะห่างของคำบริบทจากคำเป้าหมายว่าจะใช้กี่คำ ในงานวิจัยนี้ ผู้วิจัยกำหนดกรอบหน้าต่างเป็น 2 ตำแหน่ง 3 ตำแหน่ง และ 5 ตำแหน่ง มาประมวลผล เช่น การดึงคำบริบท 3 คำทางด้านซ้ายของคำเป้าหมาย

อาวุโส|กว่า|ขุนนาง|ทุกคนใน|ที่|นั้น|พลัน|ประกาศ|เสียง|กร้าว|เพื่อ|ยุติ|คำ|วิพากษ์|อัน|ฟัง

จากตัวอย่าง จะได้คำบริบทมา 3 คำ ได้แก่ *นั้น* *พลัน* *ประกาศ* ซึ่งจะดึงคำบริบทออกมาเช่นนี้จากทุกตัวอย่างข้อมูล เมื่อได้คำบริบทออกมาทั้งหมดแล้ว ก็จะนำมาหาค่าความถี่ของการปรากฏร่วมกัน จากตัวอย่างข้างต้นจะนับความถี่ได้ดังนี้ *นั้น-พลัน* ปรากฏร่วมกัน 1 ครั้ง *นั้น-ประกาศ* 1 ครั้ง และ *พลัน-ประกาศ* 1 ครั้ง และเนื่องจากในงานวิจัยนี้สนใจเรื่องลำดับของคำปรากฏรวมด้วย ดังนั้นจึงจะไม่มีการนับความถี่ในกรณีที่คำสลับตำแหน่งกัน เช่น *พลัน-นั้น*, *ประกาศ-นั้น*, *ประกาศ-พลัน* ทั้ง 3 คู่คำจึงมีค่าเท่ากับ 0 ดังแสดงไว้ในตารางที่ 3.8

ตารางที่ 3.8 การนับความถี่ของคำปรากฏรวม

	นั้น	พลัน	ประกาศ	ทั้ง	ไว้	แต่
นั้น	0	1	1	0	0	0
พลัน	0	0	1	0	0	0
ประกาศ	0	0	0	0	0	0
ทั้ง	0	0	0	0	2	1
ไว้	0	0	0	0	0	1
แต่	0	0	0	0	0	0

เมื่อได้จำนวนความถี่และคู่คำปรากฏพร้อมทั้งหมดแล้ว ค่าทั้งหมดจะนำมาแสดงอยู่ในรูปของเมทริกซ์ที่มีจำนวนแถวและคอลัมน์เท่ากัน จากนั้นจะแปลงค่าความถี่ของข้อมูลเป็นค่า log-likelihood (Dunning, 1993) เนื่องจากจำนวนความถี่ที่นับมานั้นเป็นข้อมูลดิบจะยังไม่เห็นความสัมพันธ์ของคำที่ปรากฏร่วมกันมากนัก ทั้งนี้ค่าบางค่าแม้ปรากฏร่วมกันน้อยครั้ง แต่มีความสัมพันธ์กันมาก เพราะปรากฏร่วมกันตลอด ในขณะที่บางค่าแม้ปรากฏร่วมกันมาก แต่มีความสัมพันธ์กันน้อย เนื่องจากไปปรากฏร่วมกับคำอื่นมากเช่นกัน เช่น จากข้อมูล *กำหนดการ-ผลิต* ปรากฏร่วมกัน 1 ครั้ง *ไป-คลุม* ปรากฏร่วมกัน 1 ครั้ง แต่เมื่อแปลงเป็นค่า log-likelihood แล้วปรากฏว่า ค่า log-likelihood ของ *กำหนดการ-ผลิต* มีค่าเป็น 10.08 ในขณะที่ *ไป-คลุม* มีค่าเพียง 0.07 จะเห็นว่าค่า log-likelihood แตกต่างกันมากทั้งที่มีค่าความถี่เท่ากันคือ 1 ครั้ง เนื่องจาก *กำหนดการ* ไม่ได้ไปปรากฏร่วมกับคำอื่นมากนัก ดังนั้นค่า log-likelihood จึงสูง แต่ *ไป* และ *คลุม* จะปรากฏร่วมกับคำอื่นด้วย เช่น *ไป-กิน*, *ไป-มา*, *ไป-จาก*, *ของ-คลุม*, *เปลือก-คลุม* เป็นต้น ดังนั้นค่า log-likelihood ของ *ไป-คลุม* จึงต่ำ เนื่องจากทั้งสองคำมีความสัมพันธ์กันน้อย

เมื่อแปลงค่าทั้งหมดเป็นค่า log-likelihood แล้ว จากนั้นจะนำเมทริกซ์ไปลดขนาดลงด้วยวิธีการทางคณิตศาสตร์ Singular Value Decomposition (SVD) ซึ่งการลดขนาดเมทริกซ์นี้ ถือเป็นส่วนสำคัญของ LSA เพราะ LSA เชื่อว่าจะมีข้อมูลบางส่วนที่จัดว่าเป็น noise หรือส่วนที่ไม่จำเป็น ดังนั้นการลดขนาดเมทริกซ์จึงเป็นการตัดเอาส่วนที่ไม่จำเป็นออกไป เหลือไว้เพียงข้อมูลที่สำคัญเท่านั้น วิธีการลดขนาดของเมทริกซ์ของ SVD ใช้วิธีการแยกองค์ประกอบของเมทริกซ์จากเมทริกซ์เดียวให้เป็น 3 เมทริกซ์ ได้แก่ เมทริกซ์ของแถวแนวนอน เมทริกซ์ของคอลัมน์แนวตั้ง และเมทริกซ์แนวทแยง จากนั้นจะลดขนาดของเมทริกซ์ทั้ง 3 เมทริกซ์ สำหรับในงานวิจัยนี้จะลดขนาดของเมทริกซ์ให้เหลือเพียง 10% ของข้อมูลทั้งหมด สาเหตุที่เลือกใช้ข้อมูล 10% เนื่องจากผู้วิจัยได้ทดลองลดขนาดของข้อมูลให้เหลือ 5% 10% และ 15% เพื่อเปรียบเทียบผลที่ได้จากการประมวลผลข้อมูลโดยบริบทที่ใช้จะเป็นบริบททางซ้าย 2 ตำแหน่ง 3 ตำแหน่ง และ 5 ตำแหน่ง ผลที่ได้พบว่าเมื่อใช้ปริมาณข้อมูลที่ต่างกัน ค่าความถูกต้องที่ได้ไม่ต่างกันมากนัก แต่ค่าความถูกต้องของข้อมูลที่ 10% จะให้ผลได้ดีกว่าเมื่อใช้กับระบบของคำที่มีนัยความหมายมาก ดังนั้นผู้วิจัยจึงเลือกใช้ข้อมูลที่ 10% อย่างไรก็ตาม การพิจารณาว่าจะลดขนาดของข้อมูลเท่าใดจึงจะเหมาะสมนั้นค่อนข้างจะขึ้นอยู่กับข้อมูลที่ใช้เป็นหลัก เนื่องจากจะต้องพิจารณาจากเมทริกซ์ที่ SVD ได้แยกองค์ประกอบแล้วร่วมด้วย ซึ่งการลดขนาดของข้อมูลก็มีผลต่อประสิทธิภาพของระบบเช่นกันดังจะเห็นได้จากการทดลองในเบื้องต้นของผู้วิจัยที่ค่าความถูกต้องของการลดข้อมูลเป็น 5% 10% และ 15% ไม่เท่ากัน

เมื่อลดขนาดแล้วจากนั้นนำค่าของแต่ละเมทริกซ์มารวมเข้าด้วยกันได้เป็นเมทริกซ์ใหม่ ซึ่งค่าใหม่นี้ SVD ได้คำนวณจากการปรากฏร่วมกันของคำกับคำอื่น ๆ เพื่อให้เห็นว่าคำ ๆ นี้มีความสัมพันธ์

กับคำอื่นอย่างไร ซึ่งลักษณะเช่นนี้คือการแสดงความสัมพันธ์ที่ซ่อนอยู่ของ LSA จากนั้นคำนวณหาค่ากลาง (centroid) ซึ่งเป็นค่าเฉลี่ยของกลุ่มคำบริบทในแต่ละตัวอย่าง (Tan et al., 2005) เมื่อได้ค่า centroid ทั้งหมดแล้ว จะหาค่าความคล้ายคลึงกันของแต่ละตัวอย่างเพื่อจัดกลุ่มความหมาย โดยใช้ K-means clustering algorithm (Lloyd, 1982) โดยวิธีการนี้จะให้เรากำหนดจำนวนขึ้นมา K กลุ่มในที่นี้คือ K ความหมาย ระบบจะคำนวณหาจุดศูนย์กลางของแต่ละกลุ่ม แล้วจากนั้นจะจัดให้ค่าของตัวอย่างที่ใกล้จุดศูนย์กลางของกลุ่มไหนมากที่สุดเข้าอยู่ในกลุ่มนั้น ในงานวิจัยนี้ กำหนดให้ค่า K ของเสียง และ บอกร เท่ากับ 3 และค่า K ของ หัว และ คิด เท่ากับ 9 ตามจำนวนความหมายที่ใช้ในงานวิจัยนี้ ซึ่งรายละเอียดการวิเคราะห์ความหมายจะกล่าวถึงในบทถัดไป

3.4 วิธีการประเมินประสิทธิภาพของระบบ

เนื่องจากในงานวิจัยนี้ เป็นการให้เครื่องจัดกลุ่มความหมายของคำโดยไม่ได้ให้เครื่องเรียนรู้ก่อน ดังนั้นข้อมูลทั้งหมดจึงเปรียบเสมือนเป็นข้อมูลทดสอบ โดยในการวัดผลจะนำคำตอบที่ได้จากเครื่องไปเทียบกับคำตอบที่ผู้วิจัยได้กำกับไว้ก่อนหน้า ซึ่งในการเทียบความถูกต้องจะไม่ใช่เป็นการเทียบกันโดยตรง กล่าวคือ ในการจับกลุ่มทางความหมายของเครื่อง กลุ่มที่ 1 ของเครื่องไม่จำเป็นว่าต้องเป็นความหมายที่ 1 ที่ผู้วิจัยกำกับไว้เสมอไป ความหมายที่ 1 ของผู้วิจัยอาจเป็นกลุ่มใดก็ได้ ขึ้นอยู่กับว่าตัวอย่างข้อมูลในกลุ่มนั้น ๆ ตรงกับคำตอบที่ 1 ของผู้วิจัยมากน้อยแค่ไหน ซึ่งในการกำหนดว่ากลุ่มใดเป็นคำตอบไหนนั้น ผู้วิจัยได้ใช้โปรแกรมในการคำนวณการเรียงสับเปลี่ยน (permutation) หาว่ารูปแบบคำตอบใดมีความถูกต้องมากที่สุด โดยโปรแกรมจะทดสอบจับคู่กลุ่มที่ 1 กับความหมายที่ 1 กลุ่มที่ 2 กับความหมายที่ 2 ไปเรื่อย ๆ แล้วจากนั้นจะทดสอบสลับกลุ่มกับความหมาย เช่น กลุ่มที่ 1 เป็นความหมายที่ 2 กลุ่มที่ 2 เป็นความหมายที่ 1 เพื่อหาว่าการจับคู่กลุ่มกับความหมายแบบใด เมื่อรวมกันแล้วมีค่าความถูกต้องมากที่สุด ก็จะให้รูปแบบนั้นเป็นคำตอบของระบบที่ใช้ทดสอบ

โดยสรุปแล้ว งานวิจัยนี้เลือกใช้วิธีการให้เครื่องจับกลุ่มความหมายของคำโดยไม่ได้ให้เครื่องเรียนรู้มาก่อน โดยข้อมูลตัวอย่างของคำแต่ละคำที่ใช้จะตัดส่วนที่ไม่จำเป็นออก เช่น เครื่องหมายต่าง ๆ ตัวเลข ช่องว่าง เพื่อให้เหลือเฉพาะข้อมูลส่วนที่เป็นคำจริง ๆ ซึ่งจะมีส่วนช่วยให้ระบบเห็นคำที่ปรากฏร่วมกับคำเป้าหมายได้ชัดเจนมากขึ้น

บทที่ 4

การวิเคราะห์ตามแนวภาษาศาสตร์

ในส่วนนี้ ผู้วิจัยจะกล่าวถึงการวิเคราะห์ความหมายของคำที่ใช้ศึกษาในงานวิจัยนี้ โดยในการวิเคราะห์ความหมาย ผู้วิจัยจะดึงข้อมูลจากคลังข้อมูลภาษาไทยแห่งชาติออกมาคราวละ 200 ตัวอย่าง และวิเคราะห์ความหมายจนกว่าจะไม่เจอความหมายใหม่ในข้อมูลที่ดึงออกมา 2 ครั้งติดกัน ในส่วนของทฤษฎี ผู้วิจัยใช้ทฤษฎีคำหลายความหมายอย่างมีหลักการ (principled polysemy) ของ Vyvyan Evans และ Andrea Tyler (2003) มาช่วยในการวิเคราะห์ ซึ่งผลที่ได้พบว่าคำที่มีความหมายน้อย คือ *เสียง* และ *บอก* จะพบความหมายตรงกับพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 ในขณะที่คำที่มีความหมายมาก คือ *หัว* และ *ติด* จะพบความหมายใหม่เพิ่มจากพจนานุกรม

ในการนำเสนอผลการวิเคราะห์ในบทนี้ ผู้วิจัยจะนำเสนอผลของความหมายที่ได้เลยซึ่งเป็นสาระสำคัญของการนำผลที่ได้ไปใช้จัดกลุ่มข้อมูลตามความหมายและทดลองต่อไป แต่เพื่อให้ผู้ที่สนใจการการวิเคราะห์ความหมายของคำตามทฤษฎีคำหลายความหมายอย่างมีหลักการนี้ได้เข้าใจแนวทางการวิเคราะห์ ผู้วิจัยจะแสดงรายละเอียดวิธีการวิเคราะห์ความหมายของ *หัว* ตามแนวทางนี้ มาหนึ่งความหมายโดยเลือกความหมายใหม่ที่เพิ่มมาจากพจนานุกรม

4.1 ความหมายของ *เสียง*

เมื่อวิเคราะห์ความหมายของ *เสียง* จากตัวอย่างข้อมูลแล้ว พบว่า *เสียง* ในงานวิจัยนี้มีความหมายตรงกับความหมายที่พบในพจนานุกรม โดยมีทั้งหมด 3 ความหมาย ได้แก่

ความหมายแรก หมายถึง “สิ่งที่รับรู้ได้ด้วยหู” เช่น

|ข้า|อากาศ|ก็|หลัว|สะ|ลึม| |ทว่า|**เสียง**|รถ|ที่|วิ่ง|เข้า|วิ่ง|ออก| |ถี่| |ๆ|

|ท่านหญิง|ได้ยิน|**เสียง**|หัวเราะ|ของ|ใคร|บาง|คน|ดัง|ขึ้น|ไม่|ไกล|นัก|

จากตัวอย่างจะเห็นว่าทั้ง *เสียงรถ* หรือ *เสียงหัวเราะ* ต่างเป็นเสียงที่ได้ยินในขณะที่เกิดเหตุการณ์นั้น ๆ ดังตัวอย่างที่มีรถวิ่งเข้าวิ่งออก ดังนั้นจึงเกิดเป็นเสียงรถให้เราได้ยิน หรือในตัวอย่างที่สอง ที่ในเหตุการณ์มีคนหัวเราะจึงเกิดเป็นเสียงหัวเราะให้เราได้ยิน ซึ่งตัวอย่างข้อมูลส่วนใหญ่ที่พบจะอยู่ในความหมายนี้

ความหมายที่สอง หมายถึง “ความเห็น” เช่น

|เขา|เป็น|ผู้นำ|ที่|ผู้ใหญ่|ต้อง|ฟัง|**เสียง**| |เป็น|คน|ซื่อสัตย์|ต่อ|อาชีพ|และ|ตัวเอง|

|เล่า|ให้|ฟัง|ว่า|บริษัท|ตอบสนอง|ต่อ|**เสียง**|เรียกร้อง|ถึง|ความ|รับผิดชอบ|ของ|บริษัท|

จากตัวอย่าง จะเห็นว่า **เสียง** ในที่นี้ คือความคิดเห็น ซึ่งจะต่างจากความหมายแรก ที่เสียงนั้นไม่จำเป็นต้องเกิดขึ้นหรือได้ยินในขณะที่เกิดเหตุการณ์นั้น ๆ เช่นในตัวอย่างแรก **เขา** และ **ผู้ใหญ่** ไม่จำเป็นต้องอยู่ในบริบทที่มีการสนทนากันจริง ๆ และสิ่งที่ผู้ใหญ่ต้องฟังในที่นี้ ไม่ได้หมายถึง **เสียง** ที่ **เขา** เปล่งออกมา หากแต่เป็นความคิดเห็นของ **เขา** เป็นต้น

ความหมายที่สาม หมายถึง “ความนิยม คະแนนเสียง” เช่น

ถ้า|ประชาชน|ผู้|มี|สิทธิ|เลือกตั้ง|โดย|**เสียง**|ข้าง|มาก|ของ|ผู้|มา|ออกเสียง|ประชามติ|

นายกรัฐมนตรี|กล่าวว่า|เรา|มอง|จำนวน|**เสียง**|เป็น|ส่วน|หนึ่ง|ของ|ข้อ|ห่วง|ใหญ่|

จากตัวอย่าง **เสียง** ในที่นี้คือคะแนนที่สามารถนับได้ ดังจะเห็นได้จากคำที่ปรากฏร่วมกับ **เสียง** คือ **ข้างมาก** หรือ **จำนวน** เป็นต้น

สรุปจำนวนข้อมูลตัวอย่างของแต่ละความหมายได้แสดงไว้ในตารางที่ 4.1 ดังนี้

ตารางที่ 4.1 จำนวนข้อมูล **เสียง**

ความหมาย	จำนวน
1. สิ่งที่ได้รับรู้ได้ด้วยหู	1874
2. ความเห็น	53
3. คະแนนเสียง	91
รวม	2018

4.2 ความหมายของ **บอก**

ในพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554 ให้ความหมายของ **บอก** ไว้ 4 ความหมายด้วยกัน คือ

1. พูดให้รู้, เล่าให้ฟัง
2. บ่งให้รู้
3. แนะนำ, สอน
4. รายงาน

ในงานวิจัยนี้ *บอก* มีความหมายอยู่ 3 ความหมายด้วยกัน คือ ความหมายที่ 1 ถึง 3 ตามพจนานุกรม สำหรับความหมายที่ 4 คือ รายงาน ผู้วิจัยจัดให้อยู่รวมกับความหมายที่ 1 เนื่องจากความหมายที่ 4 ต่างจากความหมายที่ 1 ไม่มาก กล่าวคือ ในความหมายที่ 4 มีลักษณะเป็นการพูดหรือแจ้งเรื่องราวให้รู้เช่นเดียวกับความหมายที่ 1 เพียงแต่เรื่องที่แจ้งนั้นจะมีลักษณะเป็นทางการมากกว่า ดังนั้นเมื่อความหมายที่ 1 และ 4 ต่างกันไม่มาก อีกทั้ง ทั้ง 2 ความหมายไม่ได้มีรูปแบบประโยคเฉพาะที่ต่างจากความหมายอื่น ผู้วิจัยจึงรวมความหมายที่ 1 และ 4 ไว้ด้วยกัน เหลือเพียง 3 ความหมาย

ตัวอย่างของข้อมูล ในความหมายที่ 1 เช่น

มุก|เลี้ยง|ที่|จะ|ไม่|**บอก**|ว่า|นี่|นะ|เป็น|สาขา|ที่|ฉัน|เ|ิง|ที่|สุด|เลย

จะ|แฉ|เรื่อง|เงิน|บริจาค|ตน|เลย|**บอก**|ว่า|จะ|ทำ|ใคร|ให้|ดู|หน้า|หน่อย|

จากตัวอย่าง จะเห็นว่า *บอก* ทั้งในตัวอย่างที่ 1 และ 2 คือการพูดให้รู้ เช่น ในตัวอย่างที่ 1 คือ คนอื่นไม่รู้วามุกไม่เก่งสาขานี้ และมุกก็เลี้ยงที่จะไม่พูดให้คนอื่นได้รู้ เป็นต้น ตัวอย่างข้อมูลส่วนใหญ่จะอยู่ในความหมายนี้

ความหมายที่ 2 ของบอก หรือ “บ่งให้รู้” เมื่อได้วิเคราะห์ความหมายจากข้อมูลตัวอย่างแล้วพบว่า ประธานที่เกิดร่วมกริยา *บอก* ในความหมายนี้ ส่วนใหญ่มักเป็นสิ่งที่ไม่มีชีวิต เช่น

ไม่|ไอ้|สี|น้ำ|ตาล|บาง|ตำแหน่ง|มี|รอย|ผุ|**บอก**|ความ|เก่า|แก่|ของ|มัน|

หน้า|ร้าน|ที่|ลื|อ|ค|กุญแจ|จาก|ด้าน|นอก|**บอก**|ว่า|ข้าง|ใน|ไม่มี|ใคร|เ|าน|ัน|

จากตัวอย่างแรก จะเห็นว่า สิ่งที่บ่งให้รู้ถึงความเก่าแก่ คือ รอยผุ และในตัวอย่างที่ 2 กุญแจที่ลือกอยู่ทำให้รู้ว่าไม่มีคนอยู่ด้านใน เป็นต้น จากตัวอย่างจะเห็นว่าสิ่งไม่มีชีวิตไม่สามารถทำกริยาบอกหรือพูดออกมาได้ตรง ๆ แต่เราสามารถอนุมานหรือตีความสิ่งที่แฝงอยู่ได้จากความรู้และประสบการณ์ที่ผ่านมา

ความหมายที่ 3 “แนะนำ หรือสอน” เช่น

ป้า|ด้ว|ง|รอบ|คอบ|พอ|ที่|จะ|**บอก**|ให้|เตรียม|เสื้อผ้า|เก่า|ๆ|ที่|จะ|เลอะ|ได้|

พี่|สาว|ยัง|**บอก**|ว่า|ใคร|ๆ|ก็|เห็น|**บอก**|เรา|ให้|เลิก|เถอะ|ก็|ไม่|ยอม|

จากตัวอย่าง จะเห็นว่าเราสามารถใช้คำว่า *แนะนำ* ไปแทนคำว่า *บอก* ได้โดยไม่เปลี่ยนความหมายของประโยค เป็น ป้าด้ว้งรอบคอบพอที่จะ*แนะนำ*ให้เตรียมเสื้อผ้าเก่าๆที่จะเลอะได้ หรือพี่สาวยัง*บอก*ว่าใครๆก็เห็น*บอก*เราให้เลิกเถอะ ในขณะที่ตัวอย่างในความหมายที่ 1 และ 2 เราไม่

สามารถนำคำว่า *แนะนำ* ไปใช้แทนที่คำว่า *บอก* ได้ นอกจากนี้โครงสร้างของการปรากฏร่วมที่พบบ่อยในความหมายนี้คือ

บอก + (คำนาม/สรรพนาม) + ให้

คำนามที่อยู่ระหว่าง *บอก* และ *ให้* สามารถมีหรือไม่มีก็ได้ โดยที่คำนามนี้มักเป็นบุคคลหรือชื่อบุคคล

สรุปจำนวนข้อมูลตัวอย่างของแต่ละความหมายของ *บอก* ได้แสดงไว้ในตารางที่ 4.2 ดังนี้

ตารางที่ 4.2 จำนวนข้อมูลของ *บอก*

ความหมาย	จำนวน
1. พูดให้รู้, เล่าให้ฟัง	1639
2. บ่งให้รู้	168
3. แนะนำ, สอน	211
รวม	2018

4.3 ความหมายของ *หัว*

หัว ในพจนานุกรมจะมีอยู่ทั้งหมด 12 ความหมายด้วยกัน สำหรับในงานวิจัยนี้พบความหมายของ *หัว* 12 ความหมาย โดยที่เป็นความหมายตามพจนานุกรม 8 ความหมาย และเป็นความหมายเพิ่มจากในพจนานุกรมอีก 4 ความหมาย และมีบางความหมายในพจนานุกรมที่ผู้วิจัยไม่พบในตัวอย่างข้อมูล ดังแสดงไว้ในตารางที่ 4.3 และเพื่อให้เห็นแนวการวิเคราะห์ความหมายในงานวิจัยนี้ ผู้วิจัยจะแสดงการวิเคราะห์ความหมายตามทฤษฎีคำหลายความหมายอย่างมีหลักการโดยละเอียดในความหมายที่ 13 ของ *หัว* ตามตารางที่ 4.3 ซึ่งเป็นความหมายที่ไม่ได้ให้ไว้ในพจนานุกรม

ตารางที่ 4.3 ความหมายของ *หัว* จากพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554 และจากงานวิจัยนี้

ความหมาย	พจนานุกรม	งานวิจัยนี้
1. ส่วนบนสุดของร่างกายของคนหรือสัตว์	✓	✓
2. ส่วนของพืชพันธุ์บางอย่างตอนที่อยู่ใต้ดิน	✓	✓
3. ส่วนแห่งสิ่งของบางอย่างที่อยู่ข้างหน้า หรือข้างต้น หรือแรกเริ่ม เช่น หัวเรือ หัวถนน	✓	✓

4. ส่วนที่อยู่ตรงข้ามกับหางหรือท้าย เช่น หัวแถว หัวเรือ	✓	✓
5. ส่วนแห่งสิ่งของที่เป็นยอด เช่น หัวผี, ส่วนแห่งสิ่งของที่ยื่นเด่นออกไป เช่น หัวแหลม หัวสะพาน	✓	✓
6. ในการเล่นปั้นแปะหรือโยนหัวโยนก้อย เรียกสมมุติ ด้านหนึ่งของเงินปลีกว่า ด้านหัว คู่กับ ด้านก้อย	✓	✗
7. ส่วนที่เป็นแก่นสาร เช่น หัวยา หัวเหล้า	✓	✗
8. ส่วนเริ่มต้นที่เป็นวงของตัวหนังสือ	✓	✗
9. ช่วงแรกเริ่มของเวลา เช่น หัวปี หัววัน หัวค่ำ	✓	✗
10. สติปัญญา, ความสามารถพิเศษ, ความคิดริเริ่ม เช่น เด็กคนนี้มีหัวทางดนตรี	✓	✓
11. ผู้ที่มีความคิดหนักไปทางใดทางหนึ่ง เช่น หัวกฎหมาย	✓	✓
12. ปัญญา, ความคิด เช่น หัวดี, หัวไว	✓	✓
13. ผม	✗	✓
14. นามนัยแทนสิ่งที่กล่าวถึงทั้งหมด	✗	✓
15. ข้อความสำคัญของเรื่องหรือข่าว	✗	✓
16. ส่วนของอุปกรณ์หรือเครื่องมือ	✗	✓

จากตาราง ความหมายที่ 1-9 คือ ความหมายของ หัว (1) และความหมายที่ 10-12 คือ ความหมายของ หัว (2) ในพจนานุกรม ความหมายที่ 6-9 เป็นความหมายที่ผู้วิจัยไม่พบในงานวิจัยนี้ และความหมายที่ 13-16 คือความหมายเพิ่มเติมที่ได้จากการวิเคราะห์ตัวอย่างข้อมูลโดยผู้วิจัย สำหรับความหมายที่ 3 และ 4 ของ หัว ผู้วิจัยได้รวมเข้าไว้เป็นหนึ่งความหมาย เนื่องจากความหมายทั้งสองมีความใกล้เคียงกัน และมีส่วนที่ซ้ำซ้อนกันอยู่ กล่าวคือ ในความหมายที่ 3 ที่ว่า “ส่วนแห่งสิ่งของบางอย่างที่อยู่ข้างหน้า หรือข้างต้น หรือแรกเริ่ม เช่น หัวเรือ หัวถนน” ในความเป็นจริงเมื่อกล่าวถึงส่วนที่อยู่ข้างหน้าหรือข้างต้นย่อมสามารถอนุมานหรือตีความได้ว่าเป็นส่วนที่อยู่ตรงข้ามกับส่วนหางหรือส่วนท้ายอยู่แล้ว จึงไม่ควรแยกเป็นอีกความหมายหนึ่ง เพราะจะทำให้เกิดความซ้ำซ้อนกัน ดังจะเห็นได้จากตัวอย่างที่ยกมาของความหมายที่ 3 และ 4 ในพจนานุกรมเป็นตัวอย่างเดียวกัน คือ หัวเรือ ด้วยเหตุผลดังกล่าว ผู้วิจัยจึงรวมความหมายที่ 3 และ 4 เข้าไว้ด้วยกัน ดังนั้นเมื่อรวมความหมายที่ 3 และ 4 แล้ว ความหมายของ หัว ในงานวิจัยนี้จึงจะเหลือ 11 ความหมาย

สำหรับจำนวนความหมายในงานวิจัยนี้จะต่างจากงานของ Kanokrattananukul (2001) ที่สามารถวิเคราะห์ความหมายของ หัว ได้ทั้งหมด 20 ความหมาย ทั้งนี้อาจเนื่องมาจากคลังข้อมูลที่ใช้ไม่ใช่คลังข้อมูลเดียวกัน โดย Kanokrattananukul ใช้คลังข้อมูลข่าวหนังสือพิมพ์กรุงเทพมหานคร ในขณะที่ผู้วิจัยใช้คลังข้อมูลภาษาไทยแห่งชาติ และขั้นตอนการวิเคราะห์ความหมายก็ต่างกัน โดย Kanokrattananukul ใช้วิธีการดึงข้อมูลออกมาตามจำนวนที่ต้องการก่อน แล้วค่อยมาวิเคราะห์ความหมาย แต่ผู้วิจัยใช้วิธีการดึงข้อมูลออกมาคราวละ 200 ตัวอย่าง จึงทำให้ไม่พบความหมายของ หัว ที่มีตัวอย่างข้อมูลน้อย เช่น ความหมายว่า “ด้านหนึ่งของเงินปลีก หรือด้านหัวคู่กับด้านก้อย” ซึ่งเป็นความหมายที่ 6 ในพจนานุกรม Kanokrattananukul พบเพียง 8 ตัวอย่างเท่านั้น ซึ่งในขั้นตอนการวิเคราะห์ความหมายของผู้วิจัยได้กำหนดไว้ว่าจะวิเคราะห์ความหมายจนกว่าจะไม่เจอความหมายใหม่ในข้อมูลที่ดึงออกมา 2 ครั้งติดกัน ดังนั้นแม้ภายหลังจะเจอความหมายนี้อยู่บ้างก็จะไม่นำมารวมไว้ด้วย

ตัวอย่างของ หัว ที่พบในความหมายแรกหรือ “ส่วนบนสุดของร่างกายของคนหรือสัตว์” เช่น

|เมื่อ|จะ|จะ|เข้า|ทำ|อย่าง|นั้น|ลิง|ก็|กระโดด|เหยียบ|หัว|และ|โจน|ข้าม|ไป|

|ก่อน|คดี|พิสดาร|ฆ่า|หุบ|หัว|เหยื่อ| |6| |ราย|ใน|ชั่ว|เวลา|ไม่|ถึง| |2| |เดือน|

จากตัวอย่าง หัว นี้คืออวัยวะส่วนหนึ่งของร่างกาย โดยในที่นี้คือ หัวของจระเข้ และหัวของคน ตามลำดับ

หัว ในความหมายที่ 2 หมายถึง “ส่วนของพืชพันธุ์บางอย่างตอนที่อยู่ใต้ดิน” เช่น

|เขา|เวอร์|ถึงขนาด|เอา|หัว|ทิวลิป|มา|เกี่ยว|ไข่|กิน|แทน|ที่|จะ|ใส่|หอม|หัว|ใหญ่|

|เรา|มา|ทำ|ความ|รู้จัก|ต้น|โพล|และ|หัว|โพล|กัน|ก่อน|นะ|ครับ|

ตัวอย่างของ หัว ในความหมายที่ 3 หรือ “ส่วนแห่งสิ่งของบางอย่างที่อยู่ข้างหน้า หรือ ข้างต้น” เช่น

|ก่อน|จะ|หยี|ตา|ขึ้น|มอง|นาฬิกา|ปลุก|ที่|หัว|เตียง| |ซึ่ง|บัดนี้|บอก|เวลา|หก|โมง|เช้า|

|Warning| |ตัว|บ่อ|เริ่ม|ที่|หัว|กระดาด|ต้นฉบับ|ของ|ผม|

จากตัวอย่างแรก หัวเตียง คือส่วนที่เป็นด้านที่เราจะใช้วางหมอนหนุนศีรษะ ซึ่งจะตรงข้ามกับส่วนที่เป็นปลายเตียง และ หัว ในตัวอย่างที่ 2 คือ ส่วนที่เป็นจุดเริ่มต้นในการเขียนหรือพิมพ์ข้อความ

หัว ในความหมายที่ 4 คือ “ส่วนแห่งสิ่งของที่เป็นยอด หรือส่วนแห่งสิ่งของที่ยื่นเด่นออกไป”
เช่น

|ถ้า|ริบ|รักษา|ให้|หาย|โดย|เร็ว| |เช่น| |ไป|สะกิด|หัว|หนอง|ออก|หรือ|กิน|ยา|ทา|ยา|
|ขา|สั้น|ไป|หมด| |หัว|เข้า|กระทบ|กัน| |เอื้อม|มือ|ไป|จับ|

จากตัวอย่าง หัวหนอง จะเป็นส่วนขา ๆ ที่ยื่น ยื่นออกมาจากผิวหนัง และ หัวเข่า เป็นส่วนที่ต่อขาส่วนบนและขาส่วนล่างไว้ด้วยกัน ซึ่งเวลาที่เราเหยียดขาตรงจะเห็นว่าส่วนที่เป็นหัวเข่าจะเป็นส่วนที่นูนเด่นออกมา ดังนั้นทั้ง หัวหนอง และ หัวเข่า จึงอยู่ในความหมายนี้

ความหมายที่ 5 หัว หมายถึง “สติปัญญา, ความสามารถพิเศษ, ความคิดริเริ่ม” เช่น

|นาย|นี้|ไม่มี|หัว|ทาง|ด้าน|ศิลปะ|เอา|ซะ|เลย|นะ|
|แล้ว|หลุยส์| |คาร์|เทียร์| |ผู้|มี|หัว|ด้าน|การ|ออกแบบ| |ก็|เป็น|คน|แรก|

จากตัวอย่าง หัว ในที่นี้หมายถึงความสามารถพิเศษทางด้านศิลปะในตัวอย่างที่ 1 และความคิดริเริ่มทางด้านการออกแบบในตัวอย่างที่ 2

ความหมายที่ 6 หัว หมายถึง “ผู้ที่มีความคิดหนักไปทางใดทางหนึ่ง” เช่น

|ความ|ไม่|ไว้|วาง|ใจ|ระหว่าง|นัก|ศึกษา|หัว|เอียง|ซ้าย|กับ|ฝ่าย|บ้าน|เมือง|ยัง|จน|กัน|
|หลัง|ถูก|พวก|มุสลิม|หัว|อนุรักษ|กล่าว|หา|ว่า|ขัด|กับ|หลัก|คำ|สอน|ทาง|ศาสนา|

จะเห็นว่า หัว ในตัวอย่างแรก คือ นักศึกษาไม่ได้มีความคิดในลักษณะเป็นกลาง แต่เอนเอียงไปทางด้านที่ขัดกับแนวคิดของรัฐบาล และ หัว ในตัวอย่างที่ 2 หมายถึง กลุ่มมุสลิมที่เคร่งศาสนา

หัว ในความหมายที่ 7 หมายถึง “ปัญญาหรือความคิด” เช่น

|แต่|อาจ|จะมี|มโน|ทัศน์|อะไร|ใน|หัว|ที่|คำ|ต่าง| |ๆ|ใน|โลก|นี้|ไม่|สามารถ|อธิบาย|
|เรา|ฟัง|แล้วก็|ง|ว่า| |ฝรั่ง|นี้|มี|อยู่|ใน|หัว|เลย|นะ| |จะ|ไป|เจอ|ยัง|ไง|

หัว ในความหมายนี้หมายถึงสมองที่เราใช้จดจำสิ่งต่าง ๆ ซึ่งในบางตัวอย่างเราสามารถใช้คำว่า สมอง แทนได้ ดังตัวอย่างที่ยกมาข้างต้น เมื่อใช้คำว่า สมอง แทนคำว่า หัว แล้วไม่ทำให้ความหมายเปลี่ยน ทั้งนี้เพราะความหมายหนึ่งของ สมอง หมายถึง ปัญญาความคิด ดังนั้น หัว ในความหมายนี้กับ สมอง จึงเป็นคำพ้องความหมายกัน

ความหมายที่ 8 หัว หมายถึง “ผม” หากวิเคราะห์ตามทฤษฎีความหลายความหมายอย่างมีหลักการแล้วจะเป็นดังนี้ ผังภูมิโครงสร้างของนามวลี หัว ในความหมายว่า “ผม” ได้แก่ หัว + วิเศษณ์ เช่น

|ยอม|รับ|ไม่ได้|ที่|จะมี|คน|หัว|ดำ|อย่าง|เรา|มา|เป็น|นาย|

|ใน|รูป|ผม|ยัง|เป็น|นักเรียน|หัว|เกรียน|อยู่|เลย|

หากดูจากโครงสร้างของนามวลีในความหมายนี้ จะพบว่าคล้ายคลึงกับนามวลีทั่วไป ดังเช่น ตัวอย่างต่อไปนี้

|ชน|ชั้นนำ|และ|สามัญชน|ที่|เก่ง|หัว|ดี|บาง|คน|ก็|สามารถ|ยึด|กุม|ใช้|

|พระ|พี่|เลี้ยง|ที่|คอย|ดูแล|เจ้า|เด็ก|หัว|กลม|เลย|มี|โอกาส|ได้|มา|ร่วม|งาน|

อย่างไรก็ตาม แม้ว่าโครงสร้างของนามวลี หัว จะเหมือนกันกับความหมายอื่น แต่ด้วยเกณฑ์ด้านการอธิบายโน้ตส์ (concept elaboration criterion) คือ ความหมายนี้จะมีรายละเอียดรูปแบบของโน้ตส์ที่ต่างจากความหมายอื่นและแสดงให้เห็นได้ในระดับภาษา (linguistic level) หัว ในความหมายว่า “ผม” มักปรากฏร่วมกับคำวิเศษณ์ที่มีความหมายเกี่ยวกับสี เช่น ดำ แดง ทอง เป็นต้น หรือความหมายที่บ่งบอกให้รู้ว่าสิ่งที่คำวิเศษณ์นี้ใช้ขยายอยู่มีลักษณะเป็นเส้นที่สามารถพันกันได้ ถูกตัดให้สั้นได้ เช่น ฟุ่ ยุ่ง เกรียน เรียบ เป็นต้น และสถานการณ์ที่ผู้พูดมักใช้ หัว ในความหมายว่า “ผม” นี้มักเกิดในบริบทที่ไม่เป็นทางการ พบในภาษาพูดมากกว่าภาษาเขียน อย่างไรก็ตาม หัว ในความหมายนี้ สามารถใช้คำว่า ผม แทนได้โดยไม่ทำให้ความหมายเปลี่ยนไป เช่น คนผมดำอย่างเรา หรือนักเรียนผมเกรียน เป็นต้น

ความหมายของ หัว นี้ขยายความหมายมาจากความหมายพื้นฐานที่ว่า “ส่วนบนสุดของร่างกายของคนหรือสัตว์” โดย ผม เป็นอวัยวะหนึ่งเป็นคนที่ขึ้นอยู่บนศีรษะ จึงถือได้ว่าเป็นความหมายส่วนหนึ่งที่อยู่ในความหมายพื้นฐานนี้ กระบวนการที่ทำให้ หัว หมายถึง “ผม” เกิดจากกระบวนการนามนัยแบบเน้นส่วนประกอบย่อย (Reimer, 2005 อ้างถึงใน จิรัชย์, 2550) โดย หัว ไม่ได้หมายถึงส่วนบนสุดทั้งหมดแต่หมายถึงเฉพาะส่วนที่เป็นผมเท่านั้น กระบวนการนี้เป็นกระบวนการเปลี่ยนแปลงทางความหมายที่ทำให้ความหมายแคบลง (semantic narrowing) เป็นการทำให้ความหมายส่วนหนึ่งเด่นขึ้นมาจนกลายเป็นความหมายใหม่

หัว ในความหมายที่ 9 หมายถึง “นามนัยแทนสิ่งที่กล่าวถึงทั้งหมด” ความหมายนี้จะสลับกับความหมายที่ 8 ที่เป็นนามนัยเชื่อมโยงความหมายแบบส่วนย่อย-ส่วนใหญ่ (part-whole) ดังเช่น

|ถ้า|เป็น|ลูก|น้อง|ผู้ชาย|ทีม|ๆ| มัน|จะ|เจด|หัว|ทิง|พัน|ที|

|อย่า|ว่า|แต่|เรือ|เลย| |แม้แต่|บ้าน|ที่|คุ้ม|หัว|ยัง|ไม่มี|จะ|อยู่|

จากตัวอย่างจะเห็นว่า หัว ในที่นี้หมายถึง ตัวบุคคล เป็นการใช หัว ที่เป็นอวัยวะหนึ่งแทนส่วนอื่นที่เหลือทั้งหมด เช่น ในตัวอย่างแรก เสด หมายถึง “การขับไล่” ในที่นี้ เสดหัว ไม่ใช่หมายถึงการขับไล่เฉพาะส่วนที่เป็นหัว แต่หมายถึง การขับไล่บุคคลนั้น ๆ เป็นต้น

ความหมายที่ 10 หัว หมายถึง “ข้อความสำคัญของเรื่องหรือข่าว” มักพบในตัวอย่างประโยคที่เกี่ยวกับสื่อสิ่งพิมพ์ต่าง ๆ เช่น

|นิตยสาร|ที่|พาด|หัว|เรื่องราว|น่า|กลัว|จาก|ภัย|นิวเคลียร์|

|รายงาน|โดย|ต่าง|โปรย|หัว|ข่าว|ตรง|กัน|ว่า| |กระทรวง|มหาด| |ไทย|ของ|อังกฤษ|

จากตัวอย่างข้างต้น หัว หมายถึง ส่วนสำคัญหรือหัวเรื่องของข่าว ใช้ดึงดูดความสนใจของผู้อื่น หัว ในความหมายนี้ มักปรากฏร่วมกับคำกริยา พาด จั่ว โปรย และกับคำนามที่มีความหมายเกี่ยวกับสื่อสิ่งพิมพ์ เช่น หนังสือพิมพ์ เรื่องราว ข่าว เป็นต้น

ความหมายที่ 11 หัว หมายถึง “ส่วนของอุปกรณ์หรือเครื่องมือ” จากตัวอย่างข้อมูล จะมีข้อมูลจำนวนหนึ่งที่ หัว จะตามด้วยคำกริยาแสดงอาการหรือการกระทำ (action verb) เช่น ฉีด กรอ ดุด เผ่า จำย เป็นต้น แสดงให้เห็นว่า หัว เป็นอุปกรณ์หรือเครื่องมือบางอย่างที่ทำหน้าที่หรือใช้งานตามคำกริยาที่ตามมา เช่น

|เครื่อง|เลเซอร์| |IPL| |จะ|เป็น|หัว|จี้|ตาม|รอย|กระ|แดด|บน|ใบหน้า|

|พื้นที่|ใน|ปั๊ม|จะ|รองรับ|หัว|จ่าย|เอ็น|จีวี|ได้|หรือไม่|

จากตัวอย่าง จะเห็นว่า หัว ใน หัวจี้ และ หัวจ่าย ต่างหมายถึงชิ้นส่วนของอุปกรณ์ที่ทำหน้าที่ใช้ “จี้” และใช้ “จ่าย” โดยปกติแล้ว หัว ถือว่าเป็นอวัยวะส่วนสำคัญที่ใช้สั่งการให้อวัยวะต่าง ๆ ทำงาน ดังนั้น เมื่อนำ หัว มาใช้แทนชิ้นส่วนของอุปกรณ์ หัว จึงเปรียบเหมือนเป็นส่วนสำคัญของอุปกรณ์ที่จะบอกให้รู้ว่าอุปกรณ์นั้น ๆ ใช้เพื่อทำอะไร เช่น หัวจี้ ก็ทำให้รู้ว่าอุปกรณ์นี้ใช้สำหรับจี้อะไรบางอย่าง เป็นต้น

สรุปความหมายของคำว่า หัว และจำนวนตัวอย่างข้อมูลของแต่ละความหมายได้แสดงไว้ในตารางที่ 4.4 ดังนี้

ตารางที่ 4.4 ความหมายของ หัว ที่วิเคราะห์ได้และจำนวนตัวอย่างข้อมูล

ความหมาย	จำนวนตัวอย่าง
1. ส่วนบนสุดของร่างกายของคนหรือสัตว์	999

2. ส่วนของพืชพันธุ์บางอย่างตอนที่อยู่ที่ดิน	70
3. ส่วนแห่งสิ่งของบางอย่างที่อยู่ข้างหน้า หรือข้างต้น หรือ แรกเริ่ม หรือ ส่วนที่อยู่ตรงข้ามกับหางหรือท้าย เช่น หัวเรือ หัวถนน หัวแถว	250
4. ส่วนแห่งสิ่งของที่เป็นยอด เช่น หัวผี, ส่วนแห่งสิ่งของที่ยื่น เด่นออกไป เช่น หัวแหลม หัวสะพาน	158
5. สติปัญญา, ความสามารถพิเศษ, ความคิดริเริ่ม เช่น เด็ก คนนี้มีหัวทางดนตรี	12
6. ผู้ที่มีความคิดหนักไปทางใดทางหนึ่ง เช่น หัวกฎหมาย	56
7. ปัญญา, ความคิด เช่น หัวดี, หัวไว	189
8. ผม	87
9. นามนัยแทนสิ่งที่กล่าวถึงทั้งหมด	106
10. ข้อความสำคัญของเรื่องหรือข่าว	10
11. ส่วนของอุปกรณ์หรือเครื่องมือ	69
รวม	2006

อย่างไรก็ตาม แม้ว่าจะวิเคราะห์ความหมายออกมาได้ 11 ความหมาย แต่จากตารางจะเห็นว่ามีความหมายที่มีตัวอย่างข้อมูลน้อย คือ ความหมายที่ 5 และความหมายที่ 10 ที่มีจำนวนตัวอย่างไม่ถึง 50 ตัวอย่าง ซึ่งในการประมวลผลของระบบหากตัวอย่างข้อมูลที่ให้ระบบเรียนรู้มีน้อยเกินไป ระบบก็จะไม่สามารถแยกความแตกต่างของความหมายนี้ออกจากความหมายอื่นได้ ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงจะตัดความหมายที่ 5 และความหมายที่ 10 ออก เหลือเพียง 9 ความหมายที่จะใช้ในการประมวลผลข้อมูล และผู้วิจัยได้เพิ่มตัวอย่างข้อมูลใหม่เข้าไปทดแทนจำนวนที่ตัดออกไป ดังแสดงไว้ในตารางที่ 4.5

ตารางที่ 4.5 ความหมายของ หัว และจำนวนตัวอย่างข้อมูลที่ใช้ในงานวิจัยนี้

ความหมาย	จำนวนตัวอย่าง
1. ส่วนบนสุดของร่างกายของคนหรือสัตว์	1,013
2. ส่วนของพืชพันธุ์บางอย่างตอนที่อยู่ที่ดิน	70

3. ส่วนแห่งสิ่งของบางอย่างที่อยู่ข้างหน้า หรือข้างต้น หรือ แรกเริ่ม หรือ ส่วนที่อยู่ตรงข้ามกับหางหรือท้าย เช่น หัวเรือ หัวถนน หัวแถว	254
4. ส่วนแห่งสิ่งของที่เป็นยอด เช่น หัวผี, ส่วนแห่งสิ่งของที่ยื่น เด่นออกไป เช่น หัวแหลม หัวสะพาน	158
5. ผู้ที่มีความคิดหนักไปทางใดทางหนึ่ง เช่น หัวกฎหมาย	56
6. ปัญญา, ความคิด เช่น หัวดี, หัวไว	191
7. ผม	89
8. นามนัยแทนสิ่งที่กล่าวถึงทั้งหมด	106
9. ส่วนของอุปกรณ์หรือเครื่องมือ	69
รวม	2006

4.4 ความหมายของ ดิด

ในพจนานุกรม *ดิด* จะมีความหมายอยู่ทั้งหมด 15 ความหมายด้วยกัน แต่จากตัวอย่างข้อมูล
ผู้วิจัยวิเคราะห์ความหมายออกมาได้ 18 ความหมาย โดยเป็นความหมายตรงตามพจนานุกรม 13
ความหมาย และเป็นความหมายที่นอกเหนือจากพจนานุกรมอีก 5 ความหมาย ดังแสดงไว้ในตารางที่
4.6

ตารางที่ 4.6 ความหมายของ *ดิด* จากพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554 และจาก
งานวิจัยนี้

ความหมาย	พจนานุกรม	งานวิจัยนี้
1. อาการที่ข้องอยู่ เช่น ดิดอยู่ในกิเลส	✓	✓
2. ประดับ เช่น ดิดเหรียญ ดิดตรา	✓	✓
3. แนบอยู่ เช่น เนื้อติดมัน เนื้อติดกระดูก	✓	✓
4. ทำให้เกิดขัดอยู่หรือข้องอยู่จนไม่อาจให้เคลื่อนไปได้	✓	✓
5. ขอบอย่างขาดไม่ได้ เช่น ดิดบุหรื	✓	✓
6. อาศัยไปด้วยหรือมาด้วย เช่น ดิดรถไปด้วย	✓	✓
7. ผนึก เช่น ดิดแสมตมบี้	✓	✓
8. จุด เช่น ดิดไฟ	✓	✓
9. ต่อเนื่องอย่างใกล้ชิด เช่น ห้องติดกัน	✓	✓

10. ได้รับเชื้อโรค เช่น ติดหิด	✓	✓
11. คงอยู่หรือทำให้คงอยู่ เช่น สีสติเคลื่อน	✓	✓
12. อยู่ในที่คุมขัง เช่น ติดคุก	✓	✓
13. มีหนี้สินติดค้างกันอยู่ (ปาก)	✓	✗
14. ไม่ร่วง, ไม่หลุด (ใช้แก๊สผลไม้)	✓	✗
15. คล้าย, ใกล้ข้าง เช่น หน้าตาติดไปทางพ่อ	✓	✓
16. มีไว้, พกไว้	✗	✓
17. ผ่านเกณฑ์	✗	✓
18. เคยชิน	✗	✓
19. ไม่ยอมห่าง	✗	✓
20. ติดตั้ง	✗	✓

จากตารางที่ 4.6 ความหมายที่ 1-15 คือความหมายตามพจนานุกรม ความหมายที่ 16-20 คือความหมายที่ผู้วิจัยวิเคราะห์ได้เพิ่มเติมจากตัวอย่างข้อมูล และความหมายที่ 13-14 คือความหมายที่ไม่พบในงานวิจัยนี้

ความหมายที่ 1 ติด หมายถึง “อาการที่ข้องอยู่” เช่น

มีทั้งชีวิตที่ทำแล้วเป็นเหตุให้หลงวน|ติด|อยู่กับการเวียนว้ายตายเกิด|
 น่ารัก|น่าใคร่|ขึ้นมา|ล่อ|ตา|ล่อ|ใจ|ให้|หลง|ติด|รัก|ชนิด|นี้|อาจ|เรียก|แรง|ทะยาน|

จากตัวอย่าง จะเห็นว่า ติด ในความหมายนี้จะ เป็นลักษณะการติดอยู่ในสิ่งที่ เป็นนามธรรม เช่น การเวียนว้ายตายเกิด ในตัวอย่างที่ 1 และ ความรัก ในตัวอย่างที่ 2 นอกจากนี้ พบว่าคำที่มัก ปรากฏร่วมกับคำว่า ติด ในความหมายนี้คือคำว่า ยึด เช่น

ก็|ยัง|มี|ความ|เชื่อ|หรือ|ติด|ยึด|กับ|ค่านิยม|ที่|ว่า|นี้|ตลอด|มา|
 ข้า|คน|เป็น|ว่า|เล่น| |หลง|หลง|มง|าย|ยึด|ติด|กับ|ทฤษฎี| |มี|จิตใจ|คับ|แคบ|

ติด ในความหมายที่ 2 หมายถึง “ประดับ” เช่น

ผม|หน้า|ปัด|เป้|ติด|กับ|สี|ฟ้า|น่ารัก|
 คณะ|ครู|ที่|ได้|ติด|ชิด|ใหม่|ได้|กิน|เลี้ยง|ฉลอง|ชิด|โดย|เชิญ|หัวหน้า|หมวด|

ติด ในความหมายนี้ หมายความว่าถึง การประดับตกแต่ง รวมถึงการประดับตำแหน่งและยศ
 ดังเห็นได้จากตัวอย่างที่ 1 เป็นการใช้กับประดับผม และในตัวอย่างที่ 2 เป็นการประดับชิตหรือยศ
 ใหม่ เป็นต้น

ความหมายที่ 3 ติด หมายถึง “แนบอยู่” เช่น เนื้อติดมัน เนื้อติดกระดูก เช่น

|ชาย|คน|แรก|มี|ผม|สั้น|สี|ดำ|ขลุ่ย|หวี|เรียบ|**ติด**|หัว| |สวม|เสื้อ|แขน|ยาว|สี|ดำ|
 |และ|แม่|แต่|หมู|**ติด**|หนัง|ก็|เอา|มา|แบ่ง|ต้ม|เค็ม|ได้|

ในความหมายนี้ คือ ของ 2 สิ่งอยู่แนบชิดกันมาก จนเหมือนเป็นของชิ้นเดียวกัน ดังตัวอย่าง
 ที่ 1 ที่ ผม ถูกหวีจนเรียบแนบติดไปกับหัว หรือ หนัง ในตัวอย่างที่ 2 ที่แนบติดอยู่กับชิ้นเนื้อหมู เป็น
 ต้น

ติด ในความหมายที่ 4 หมายถึง “ทำให้เกิดขัดอยู่หรือข้องอยู่จนไม่อาจให้เคลื่อนไปได้” เช่น

|แถม|เธอ|ยัง|โง่|ร้าย|เป็น|พิเศษ|ที่|**ติด**|ไฟ|แดง|มา|เกือบ|ทุก|แยก|
 เพราะ|เห็น|คน|รับ|โทรศัพท์|บอก|ว่า|**ติด**|ประชุม|ตลอด|ก็|เลย|ขอ|เลื่อน|ลูกค้า|

ติด ในความหมายนี้ ไม่ได้หมายถึงเฉพาะการติดไฟแดงหรือการติดฝนเพียงอย่างเดียว แต่
 รวมถึงการติดขัดอยู่กับสิ่งใดสิ่งหนึ่งแล้วทำให้สิ่งอื่นดำเนินต่อไปไม่ได้ ดังเช่น ในตัวอย่างที่ 2 ที่เมื่อติด
 ประชุมแล้วก็มีผลต่อเนื่องคือทำให้ไม่สามารถไปคุยกับลูกค้าได้ จึงต้องมีการเลื่อนลูกค้าออกไป เป็น
 ต้น

ความหมายที่ 5 ติด หมายถึง “ชอบอย่างขาดไม่ได้” ติด ในความหมายนี้จะตรงข้ามกับ ติด
 ในความหมายที่ 1 กล่าวคือ ติด ในความหมายนี้จะเป็นการติดอยู่กับสิ่งที่เป็นรูปธรรมในขณะที่ใน
 ความหมายที่ 1 จะเป็นเรื่องของนามธรรมเป็นหลัก ในพจนานุกรมยกตัวอย่าง ติดผืน ติดบุหรื หรือ
 จากตัวอย่างในข้อมูล เช่น

|ใน|ที่|สุด|ลูก|คน|หนึ่ง|ก็|**ติด**|ยา|เฮโรอีน|

|รายการ|พิเศษ|แทรก|เข้ามา|เยอะ|แล้ว|คน|ที่|**ติด**|ละคร|เรื่อง| |"สาว|น้อย"| |ก็|จะ|
 บ่น|กัน|มาก|

ความหมายที่ 6 ติด หมายถึง “อาศัยไปด้วยหรือมาด้วย” เช่น ติดรถ ตามตัวอย่างใน
 พจนานุกรม หรือจากตัวอย่างข้อมูล เช่น

|น้ำที่ไหลออกมา|มีอัตราเร็ว|กว่า|ปกติ|และมี|ตะกอน|**ติด**|ออกมา|ด้วย|

|หรือ|บาง|ครั้ง|เด็ก|ที่|**ติด**|ไป|กับ|พ่อแม่|หรือ|ยาย|ยาย| |ครั้ง|ผู้ใหญ่|ไป|เล่น|การ|พนัน|

จากตัวอย่างที่ 1 คือ น้ำที่ไหลออกมามีตะกอนออกมาพร้อมกับน้ำ หรือในตัวอย่างที่ 2 คือ เด็กมาพร้อมกับพ่อแม่ ทั้งสองตัวอย่างต่างแสดงลักษณะการ “อาศัยมาด้วย” ด้วยเช่นกัน กล่าวคือ ในตัวอย่างแรก หากไม่มีน้ำที่ไหล ตะกอนก็ไม่สามารถหลุดออกมาได้ หรือในตัวอย่างที่ 2 หากไม่มีผู้ใหญ่ไปเด็กก็ไม่สามารถไปในสถานที่เล่นการพนันได้ เป็นต้น

ความหมายที่ 7 **ติด** หมายถึง “ผนึก” เช่น

|ทั้งนี้|ทาง|บริษัท|ได้|ทำ|สลาก|**ติด**|ข้าง|ขวด|ไว้|เพื่อ|บอกว่า| |นี่|เป็น|เบียร์|

|กระดาน|แผ่น|สี|เขียว|เขียน|ขอ|บริจาค|สรรพสิ่ง|ของ|**ติด**|ไว้|ทาง|บันได|ขึ้น|ลง|

จากตัวอย่าง **ติด** คือ การเอาของยึดติดกันด้วยกาวหรือด้วยสิ่งที่มีคุณสมบัติสามารถยึดติดของสิ่งหนึ่งไว้กับอีกสิ่งหนึ่งได้ เช่น เอาสลากผนึกไว้กับขวด เป็นต้น

ติด ในความหมายที่ 8 หมายถึง “จุด” เช่น ติดไฟ ติดเตา สำหรับความหมายนี้ผู้วิจัยได้รวมเอาความหมายว่า “การเปิดเครื่องหรือทำให้เครื่องจักรทำงาน” เข้าไว้ด้วย เนื่องจาก ความหมายของการจุด ก็เหมือนการทำให้ของสิ่งหนึ่งเริ่มใช้งานได้ เช่น **ติดเตา** ก็คือ การก่อไฟให้เตาสามารถใช้งานได้ **ติดตะเกียง** ก็คือการจุดไฟให้ตะเกียงสามารถใช้งานได้ และในทำนองเดียวกัน **ติดเครื่อง** ก็คือ การเริ่มงานของเครื่องยนต์ให้พร้อมใช้งานเช่นกัน ด้วยเหตุผลดังกล่าวผู้วิจัยจึงขยายความหมายของ **ติด** ในความหมายที่ 8 ให้รวมครอบคลุมถึงการทำให้สิ่งใดสิ่งหนึ่งพร้อมใช้งานนอกเหนือจากความหมายว่า “จุด” ตัวอย่างเช่น

|เขา|เปิด|ประตู|เข้าไป|ใน|รถ| |**ติด**|เครื่องยนต์|ด้วย|อาการ|รีบร้อน|

ระบบ|เซนเซอร์|จะ|สั่งการ|ให้|คอมพิวเตอร์|ที่|โซน| A| **ติด** |และ|ไนท์|ไลท์|

จากตัวอย่างที่ 1 คือ การทำให้เครื่องยนต์ทำงาน และในตัวอย่างที่ 2 คือการทำให้คอมพิวเตอร์เปิดเพื่อพร้อมใช้งาน เป็นต้น

ติด ในความหมายที่ 9 หมายถึง “ต่อเนื่องอย่างใกล้ชิด” เช่น

|ที่จะ|ไป|ถนน|เจริญ|นคร| |จะมี|ที่|จอดรถ|อยู่|**ติด**|ถนน|ใหญ่| |ไม่ต้อง|เข้า|ซอย|

|แต่|เรามี|เพื่อน|เป็น|เด็ก|ฝรั่ง|บ้าน|**ติด**| ๆ|กัน| |พ่อ|มี|บ้าน|ให้|พวก|ฝรั่ง|เช่า|สอง|

ในความหมายนี้ หมายถึง ของสองสิ่งจะอยู่ชิดเรียงต่อกัน ซึ่งคำปรากฏร่วมที่มักปรากฏร่วมกับ *ติด* ในความหมายนี้คือ ไม้มก เป็นการแสดงให้เห็นถึงความต่อเนื่องกัน ดังจะเห็นได้จากตัวอย่างที่ 2 ที่แสดงให้เห็นว่าบ้านแต่ละหลังอยู่เรียงต่อกันไปอย่างต่อเนื่อง เป็นต้น

ความหมายที่ 10 *ติด* หมายถึง “ได้รับเชื้อโรค” เช่น

|คนเรา|สามารถ|*ติด*|โรค|จาก|สัตว์|ได้|หลาย|ทาง|

|แน่นอน|ว่า|อัตรา|การ|*ติด*|เชื้อ|โรค|ระหว่าง|คน|ก็|จะ|เพิ่ม|มาก|ขึ้น|

ความหมายที่ 11 *ติด* หมายถึง “คงอยู่หรือทำให้คงอยู่” เช่น

|ของ|ใคร| |บอก|ให้|เขียน|ชื่อ|*ติด*|ไว้|ทำไม|ไม่|รู้จัก|เขียน|

|ยัง|ได้|พบ|คราบ|เลือด|ของ|นาทาล|*ติด*|อยู่|ที่|กล่อง|พิซซา|กล่อง|หนึ่ง|

ในตัวอย่างแรก คือ การทำให้ *ชื่อ* คงอยู่กับสิ่งของบางอย่าง หรือในตัวอย่างที่ 2 คือ *เลือด* ยังคงอยู่ที่กล่องพิซซา เป็นต้น

ติด ในความหมายที่ 12 หมายถึง “อยู่ในที่คุมขัง” เช่น

|สอง|ตัว|ที่|ได้|พบ|เข้า|นี้|ร้อง|เสียง|ไพเราะ|กว่า|นก|ที่|*ติด*|ขัง|อยู่|ใน|กรง|

|คนที่|ทำ|ท่า|เหมือน|จะ|ทิ้ง|ให้|เธอ|*ติด*|เกาะ|อยู่|คน|เดียว|จริง|ๆ|

ที่คุมขัง ในที่นี้ ไม่จำเป็นต้องหมายถึง ตระราชหรือคุก เท่านั้น แต่ยังสามารถหมายถึงสถานที่อื่น ๆ ที่ทำให้ขาดอิสรภาพ ถูกจำกัดบริเวณได้เช่นกัน ดังเช่น *เกาะ* ที่ยกมาในตัวอย่างที่ 2 *เกาะ* แม้จะใหญ่ และไม่มีลูกกรงกั้นดังเช่นคุก แต่ *เกาะ* มีน้ำทะเลล้อมรอบ ดังนั้น *เกาะ* จึงไม่ต่างจากสถานที่ที่คุมขัง เพราะคนที่อยู่บนเกาะจะไม่สามารถไปที่อื่นได้ หากไม่มีเรือหรือพาหนะอื่นพาออกมา

ความหมายที่ 13 *ติด* หมายถึง “คล้าย, ใกล้ข้าง” เช่น

|มอง|พี่น้อง|ทะเลาะ|กัน|ด้วย|ความรู้สึก|ที่|*ติด*|จะ|ขบขัน|ไม่น้อย|

|ไม่ได้|ยิ้ม|ให้|อย่าง|ทุก|ที| |กลับ|มี|สี|หน้า|*ติด*|ไป|ทาง|เคร่ง|เครียด|จาก|คิ้ว|ที่|ขมวด|

ในความหมายนี้ เราสามารถใช้คำว่า *คล้าย* แทน *ติด* ได้โดยไม่ทำให้ความหมายเปลี่ยน เป็น *ความรู้สึกที่คล้ายจะขบขันไม่น้อย* หรือ *มีสีหน้าคล้ายไปทางเคร่งเครียด* และคำปรากฏร่วมที่มักปรากฏร่วมกับ *ติด* ในความหมายนี้คือ *จะ* เป็น *ติดจะ* ซึ่งจะไม่พบในความหมายอื่น

ติด ในความหมายที่ 14 หมายถึง “มีไว้ พกไว้” เช่น

|ต้องมี|ผลิตภัณฑ์| |‘No.7’| |*ติด*|ใน|กระเป๋า|เครื่อง|สำอาง|อย่างน้อย| |1| |ชั้น|

|เรียก|กัน|ฉบับ|ย่อ|ว่า|ซีลี|เฉย|ๆ|มี|**ติด**|ช่อง|แข็ง|ไว้|ช่วย|ชีวิต|ลูก|ผู้|หิว|โหย|

ติด ในความหมายนี้ คือ การมีสิ่งใดสิ่งหนึ่งไว้ ส่วนใหญ่มีไว้เพื่อสำหรับใช้ประโยชน์ รูปแบบของประโยคที่มักพบในความหมายนี้ คือ

มี/เอา + (สิ่งของ) + **ติด**

ดังจะเห็นได้จากตัวอย่างที่ยกมาข้างต้น

ความหมายที่ 15 **ติด** หมายถึง “ผ่านเกณฑ์” เช่น

|สิ่ง|ไม่|คาด|ฝัน|ก็|เกิด|ขึ้น|ฉัน|สอบ|**ติด**|คณะ|โบราณคดี| มหาวิทยาลัยศิลปากร|ได้|

|ได้รับ|เลือก|ให้|**ติด**| 1| ใน| 100| อันดับ|ผู้|หญิง|ที่|แซ่|ซี้|ที่สุด|ใน|โลก|

จากตัวอย่างแรก หมายถึง สอบผ่านเกณฑ์คะแนนของมหาวิทยาลัย และในตัวอย่างที่ 2 คือผ่านเกณฑ์การคัดเลือกให้ติดอันดับโลก เป็นต้น

ความหมายที่ 16 **ติด** หมายถึง “เคยชิน” เนื่องจากได้ทำอยู่เป็นประจำ เช่น

|ที่|ปลูก|ฝัง|ถ่ายทอด|กัน|มา|จน|**ติด**|เป็น|นิสัย|ใจคอ| |ดั่ง|เช่น|เอก|ลักษณ์|ทั้งหลาย|

|น้ำ|ๆ|เคย|เปิด|ให้|ฟัง|บ่อย|จน|หล่น|จำ|ได้|**ติด**|หู|ขึ้น|มา|ใช้|ทัน|ควัน|

ในตัวอย่างแรก คือ การได้ทำบางอย่างสืบทอดกันมา จนทำให้เกิดความเคยชินจนเป็นนิสัย ส่วนตัวอย่างที่ 2 คือ การได้ฟังอยู่เป็นประจำจึงทำให้เกิดการคุ้นหู

ติด ในความหมายที่ 17 หมายถึง “ไม่ยอมห่าง” เช่น

|กลายเป็น|ไซ้|ผูก|ขา|สามี|ให้|อยู่|**ติด**|บ้าน|ได้|อีก|เหตุผล|หนึ่ง|ด้วย|

|บาง|ครั้ง|จะ|ร้อง|เรียก|และ|**ติด**|แม่|อย่าง|หนัก|หน่วง|กว่า|เมื่อ|ยัง|เป็น|ทารก|

จากตัวอย่างคือ **อยู่ติดบ้าน** คือ ไม่ค่อยออกจากบ้านไปที่อื่น หรือ **ติดแม่** คือ ไม่ยอมอยู่ห่างจากแม่ ซึ่ง **ติด** ในความหมายนี้ มักเป็นการติดบุคคล เช่น ติดเพื่อน ติดพ่อ เป็นต้น

ความหมายที่ 18 **ติด** หมายถึง “ติดตั้ง” เช่น

|เขา|มี|แผน|จะ|เปิด|โรงแรม|สุนัข|ที่|จะ|**ติด**|กล่อง|วงจร|ปิด|เพื่อให้|เจ้าของ|

|ภายใน|ห้อง|ที่|เป็น|หน้าต่าง|กระจก|**ติด**|เหล็ก|ดัด|แข็งแรง|ผิด|จาก|ห้อง|อื่นๆ|

ในความหมายนี้ สามารถใช้คำว่า **ติดตั้ง** แทนคำว่า **ติด** ได้โดยไม่ทำให้ความหมายเปลี่ยน เช่น โรงแรมสุนัขที่จะ**ติดตั้ง**กล่องวงจรปิด หน้าต่างกระจก**ติดตั้ง**เหล็กดัด เป็นต้น

สรุปความหมายของคำว่า *ติด* และจำนวนตัวอย่างข้อมูลของแต่ละความหมายได้แสดงไว้ในตารางที่ 4.7 ดังนี้

ตารางที่ 4.7 ความหมายของ *ติด* ที่วิเคราะห์ได้และจำนวนตัวอย่างข้อมูล

ความหมาย	จำนวนตัวอย่าง
1. อาการที่ซึ้งอยู่ เช่น ติดอยู่ในกิลเลส	203
2. ประดับ เช่น ติดเหรียญ ติดตรา	49
3. แนบอยู่ เช่น เนื้อติดมัน เนื้อติดกระดูก	41
4. ทำให้เกิดขัดอยู่หรือซึ้งอยู่จนไม่อาจให้เคลื่อนไปได้	293
5. ซอบอย่างขาดไม่ได้ เช่น ติดบุหรี	192
6. อาศัยไปด้วยหรือมาด้วย เช่น ติดรถไปด้วย	27
7. ผนึก เช่น ติดแสตมป์	222
8. จุด เช่น ติดไฟ	28
9. ต่อเนื่องอย่างใกล้ชิด เช่น ห้องติดกัน	417
10. ได้รับเชื้อโรค เช่น ติดหิด	296
11. คงอยู่หรือทำให้คงอยู่ เช่น สีติดเสื้อ	187
12. อยู่ในที่คุมขัง เช่น ติดคุก	57
13. คล้าย, ใกล้ข้าง เช่น หน้าตาติดไปทางพ่อ	24
14. มีไว้, พกไว้	45
15. ผ่านเกณฑ์	57
16. เคยชิน	36
17. ไม่ยอมห่าง	31
18. ติดตั้ง	16
รวม	2221

แม้ว่าจำนวนความหมายของ *ติด* ที่วิเคราะห์ได้จะมีมากถึง 18 ความหมาย แต่มีหลายความหมายด้วยกันที่มีจำนวนตัวอย่างไม่ถึง 50 ตัวอย่าง ซึ่งผู้วิจัยจะตัดความหมายเหล่านั้นออกด้วยเหตุผลเดียวกับที่ตัดความหมายของ *หั่ว ออก* ดังนั้นความหมายที่จะใช้ในงานวิจัยนี้ได้สรุปไว้ดังแสดงไว้ในตารางที่ 4.8 ดังนี้

ตารางที่ 4.8 ความหมายของ *ติด* และจำนวนตัวอย่างข้อมูลที่ใช้ในงานวิจัยนี้

ความหมาย	จำนวนตัวอย่าง
1. อาการที่ข้งอยู่ เช่น ติดอยู่ในกิเลส	213
2. ทำให้เกิดขัดอยู่หรือข้งอยู่จนไม่อาจให้เคลื่อนไปได้	303
3. ขอบอย่างขาดไม่ได้ เช่น ติดบุหรี	200
4. ผนึก เช่น ติดแสดมภ์	227
5. ต่อเนื่องอย่างใกล้ชิด เช่น ห้องติดกัน	435
6. ได้รับเชื้อโรค เช่น ติดหิด	308
7. คงอยู่หรือทำให้คงอยู่ เช่น สติดเลื้อ	201
8. อยู่ในที่คุมขัง เช่น ติดคุก	59
9. ผ่านเกณฑ์	60
รวม	2006

จากผลการวิเคราะห์ความหมายของคำนาม *เสียง หั่ว* และคำกริยา *บอก ติด* พบว่าคำกริยามี นัยความหมายหลากหลายมากกว่าคำนามตามสมมติฐาน ในคำที่มีจำนวนความหมายน้อย คือ *เสียง* และ *บอก* จะไม่เห็นความแตกต่าง เนื่องจากมีจำนวน 3 ความหมายเท่ากัน แต่จะเห็นได้ชัดในคำที่มี จำนวนความหมายมาก คือ *หั่ว* และ *ติด* โดยคำว่า *ติด* สามารถวิเคราะห์ความหมายได้มากถึง 18 ความหมาย ในขณะที่ *หั่ว* สามารถวิเคราะห์ได้เพียง 11 ความหมาย ทั้งนี้อาจเนื่องมาจากตัวอย่าง ข้อมูลส่วนใหญ่ของ *หั่ว* ไม่มีการกระจายมากนัก โดยจะอยู่ในความหมายที่ 1 (ส่วนบนสุดของร่างกาย ของคนหรือสัตว์) มากที่สุดซึ่งมีถึงครึ่งหนึ่งของข้อมูลทั้งหมด คือ 1,013 ตัวอย่างจากทั้งหมด 2,006 ตัวอย่าง ในขณะที่จำนวนตัวอย่างของ *ติด* จะค่อนข้างกระจาย โดยความหมายที่มีจำนวนตัวอย่าง มากที่สุดมีตัวอย่างข้อมูลที่ประมาณ 1 ใน 4 ของตัวอย่างทั้งหมด คือ 435 ตัวอย่าง จาก 2,006 ตัวอย่าง สำหรับคำที่มีความหมายน้อยพบว่าทั้ง *เสียง* และ *บอก* จะมีความหมายหลักที่มีตัวอย่าง ข้อมูลค่อนข้างมากเมื่อเทียบกับความหมายอื่น กล่าวคือ ประมาณ 75% ของข้อมูลจะอยู่ใน ความหมายหลัก โดยความหมายหลักของ *เสียง* มีตัวอย่างข้อมูลมากถึง 1,874 ตัวอย่างจาก 2,018 ตัวอย่าง และความหมายหลักของ *บอก* มีตัวอย่างข้อมูล 1,639 ตัวอย่างจาก 2,018 ตัวอย่าง

ในบทนี้ ผู้วิจัยได้วิเคราะห์ความหมายของคำที่ใช้ศึกษาทั้ง 4 คำ ตามแนวภาษาศาสตร์ เพื่อให้รู้จำนวนความหมายของคำที่ใช้ รวมถึงทำให้เห็นรูปแบบของคำหรือบริบทที่ปรากฏร่วมกันกับคำที่ศึกษานั้น ๆ ซึ่งจะเป็นประโยชน์ในการนำมาเปรียบเทียบกับ การแยกความหมายด้วยวิธีการทางคอมพิวเตอร์ว่าคอมพิวเตอร์จะแยกความหมายได้อย่างไร และใช้บริบทเดียวกันกับที่วิเคราะห์ได้ตามแนวภาษาศาสตร์หรือไม่ ซึ่งในบทต่อไปผู้วิจัยจะได้กล่าวถึงวิธีการแยกความหมายของคำด้วยเครื่อง

บทที่ 5

การทดลองแยกความหมายของคำด้วยเครื่อง

ในบทนี้ ผู้วิจัยจะกล่าวถึงการทดลองแยกนัยความหมายของคำด้วยเครื่องตามขั้นตอนที่ได้อธิบายไว้ในบทที่ 3 โดยใช้คำบริบทด้านซ้าย ด้านขวา ด้านซ้าย-ขวา และกรอบหน้าต่างหรือความห่างของคำปรากฏร่วมกับคำเป้าหมายในระยะต่าง ๆ เพื่อศึกษาว่าคำบริบทจะสามารถช่วยเครื่องแยกนัยความหมายของคำได้มากน้อยเพียงใด ซึ่งในงานวิจัยนี้ไม่ได้ใช้คุณสมบัติทางภาษา เช่น หน้าที่ของคำ ในการช่วยแยกนัยความหมาย จะใช้เพียงแค่รูปคำของคำปรากฏร่วมเท่านั้น ซึ่งผลที่ได้จะนำไปเปรียบเทียบกับผลจากการวิเคราะห์ตามแนวภาษาศาสตร์ว่าเครื่องสามารถแยกนัยความหมายได้โดยใช้บริบทเดียวกันกับที่วิเคราะห์ได้ตามแนวภาษาศาสตร์หรือไม่ ในบทนี้ ผู้วิจัยได้แบ่งเนื้อหาออกเป็น 4 ส่วน ได้แก่ กรอบหน้าต่างและบริบท จำนวนความหมายของคำ ประเภทของคำ และบริบทที่เครื่องใช้แยกนัยความหมายของคำที่ใช้ในงานวิจัย

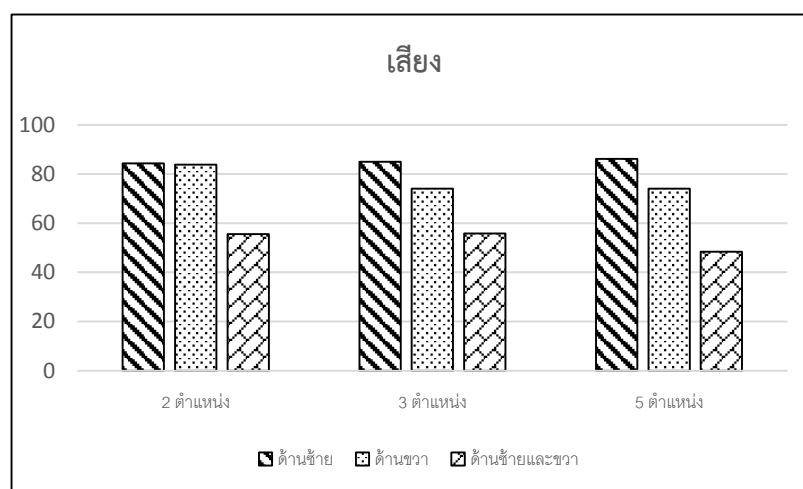
5.1 กรอบหน้าต่างและบริบท

กรอบหน้าต่างที่กล่าวถึงนี้คือระยะความห่างของคำเป้าหมายกับคำปรากฏร่วมที่เราจะกำหนดให้คำห่างออกไปเท่าใดจึงจะยังจัดว่าเป็นคำปรากฏร่วมกับคำเป้าหมายอยู่ ในงานวิจัยนี้ผู้วิจัยได้ทดลองใช้กรอบหน้าต่างที่ความห่าง 2 3 และ 5 เพื่อเปรียบเทียบประสิทธิภาพของระบบเมื่อใช้กรอบหน้าต่างที่แตกต่างกัน ซึ่งกรอบหน้าต่างนี้ได้ทดลองใช้กับคำบริบททางด้านซ้าย ด้านขวา และด้านซ้ายและขวา ซึ่งผลการทดลองเป็นดังนี้

ตารางที่ 5.1 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าต่างที่ต่างกันของคำว่า เสี่ยง

บริบท \ กรอบหน้าต่าง	2 ตำแหน่ง		3 ตำแหน่ง		5 ตำแหน่ง	
	จำนวน	%	จำนวน	%	จำนวน	%
ด้านซ้าย	1701	84.29	1714	84.94	1738	86.13
ด้านขวา	1690	83.75	1494	74.03	1494	74.03
ด้านซ้ายและขวา	1120	55.50	1126	55.80	978	48.46

หมายเหตุ : จำนวนตัวอย่างทั้งหมด 2018 ตัวอย่าง

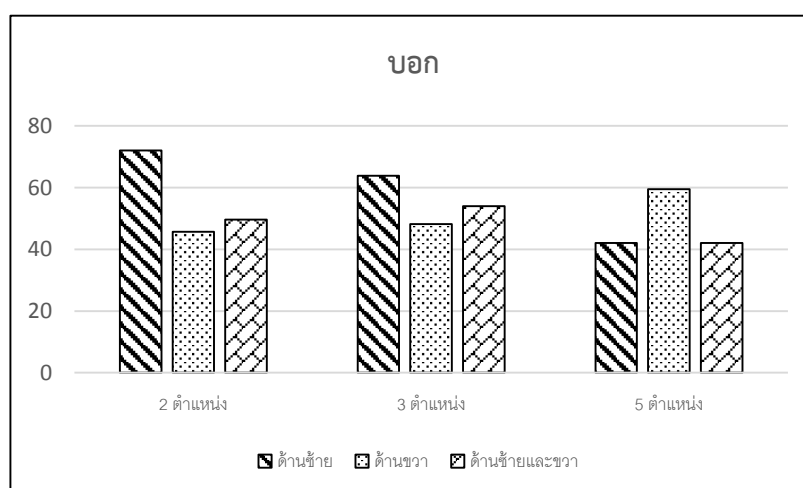


ภาพที่ 5.1 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าต่างที่แตกต่างกันของ เสียง

ตารางที่ 5.2 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าต่างที่ต่างกันของคำว่า บอก

บริบท \ กรอบหน้าต่าง	2 ตำแหน่ง		3 ตำแหน่ง		5 ตำแหน่ง	
	จำนวน	%	จำนวน	%	จำนวน	%
ด้านซ้าย	1453	72.00	1288	63.83	849	42.07
ด้านขวา	922	45.69	973	48.22	1201	59.51
ด้านซ้ายและขวา	1001	49.60	1089	53.96	849	42.07

หมายเหตุ : จำนวนตัวอย่างทั้งหมด 2018 ตัวอย่าง

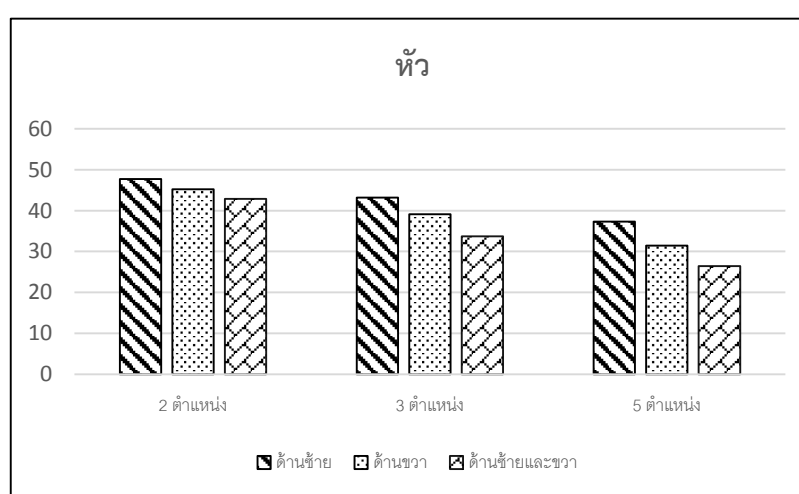


ภาพที่ 5.2 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าต่างที่แตกต่างกันของ บอก

ตารางที่ 5.3 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าที่ต่างกันของคำว่า หัว

บริบท \ กรอบหน้าที่ต่าง	2 ตำแหน่ง		3 ตำแหน่ง		5 ตำแหน่ง	
	จำนวน	%	จำนวน	%	จำนวน	%
ด้านซ้าย	958	47.76	867	43.22	749	37.34
ด้านขวา	907	45.21	785	39.13	631	31.46
ด้านซ้ายและขวา	860	42.87	676	33.70	531	26.47

หมายเหตุ : จำนวนตัวอย่างทั้งหมด 2006 ตัวอย่าง

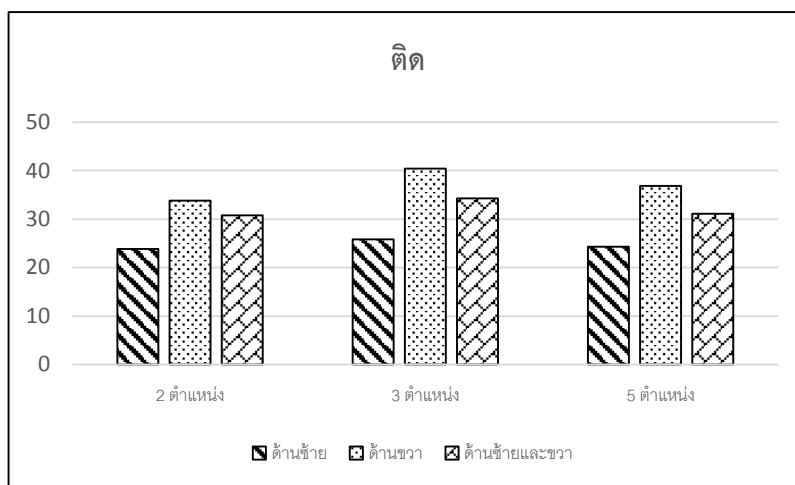


ภาพที่ 5.3 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าที่แตกต่างกันของ หัว

ตารางที่ 5.4 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าที่ต่างกันของคำว่า ติด

บริบท \ กรอบหน้าที่ต่าง	2 ตำแหน่ง		3 ตำแหน่ง		5 ตำแหน่ง	
	จำนวน	%	จำนวน	%	จำนวน	%
ด้านซ้าย	479	23.88	518	25.82	487	24.28
ด้านขวา	679	33.85	811	40.43	739	36.84
ด้านซ้ายและขวา	617	30.76	688	34.30	624	31.11

หมายเหตุ : จำนวนตัวอย่างทั้งหมด 2006 ตัวอย่าง



ภาพที่ 5.4 ค่าความถูกต้องของระบบเมื่อใช้บริบทและกรอบหน้าต่างที่แตกต่างกันของ ติด

จากผลการทดลอง พบว่าบริบททางด้านซ้ายช่วยแยกนัยความหมายได้ดีกว่าบริบทอื่นในคำว่า *เสียง* และ *หัว* ในขณะที่บริบททางขวาช่วยแยกนัยความหมายของคำว่า *ติด* ได้ดีกว่าบริบทอื่นสำหรับคำว่า *บอก* ช่วงบริบท 2 ตำแหน่งและ 3 ตำแหน่ง บริบททางซ้ายจะช่วยแยกนัยความหมายได้ดีที่สุด แต่เมื่อเปลี่ยนเป็นช่วงบริบท 5 ตำแหน่ง บริบทที่ช่วยให้ระบบสามารถแยกนัยความหมายได้ดีที่สุดกลับเป็นบริบททางขวา จากภาพกราฟของค่าทั้ง 4 ค่าจะเห็นว่าเมื่อใช้กรอบหน้าต่างที่มากขึ้น ส่วนใหญ่ค่าความถูกต้องจะลดลง จะมีเพียงคำว่า *เสียง* ที่เมื่อใช้บริบททางซ้ายแล้วค่าความถูกต้องเพิ่มขึ้นกับคำว่า *บอก* ที่เมื่อใช้บริบททางขวาแล้วค่าความถูกต้องเพิ่มขึ้นเมื่อใช้กรอบหน้าต่างที่มากขึ้น ซึ่งผู้วิจัยจะกล่าวถึงในรายละเอียดต่อไป จากตารางที่ 5.1-5.4 ค่าที่แสดงดังกล่าวเป็นค่าที่รวมจากความหมายแล้ว เมื่อแยกรายละเอียดการจัดกลุ่มของระบบตามแต่ละความหมายของแต่ละคำจะเป็นดังตารางที่ 5.5-5.8

ในตารางที่ 5.5-5.8 นี้ จะแสดงการจัดกลุ่มของระบบด้วยวิธี K-means clustering algorithm ซึ่งค่า k ผู้วิจัยได้กำหนดตามจำนวนความหมายที่วิเคราะห์ได้ของแต่ละคำตามที่ได้อธิบายไว้ในบทที่ 4 ซึ่งค่า k สำหรับ *เสียง* และ *บอก* เท่ากับ 3 และค่า k สำหรับ *หัว* และ *ติด* เท่ากับ 9 โดยข้อมูลตามแถวหมายถึงตัวอย่างข้อมูลทั้งหมดที่ผู้วิจัยได้กำกับไว้ว่ามีความหมายตามที่ระบุในคอลัมน์แรก และข้อมูลตามคอลัมน์หมายถึง จำนวนข้อมูลในแต่ละความหมายที่ระบบจัดให้อยู่ในแต่ละกลุ่มหลังจากที่ระบบได้จัดกลุ่มตามรายละเอียดที่ปรากฏในตารางแล้ว ผู้วิจัยคัดเลือกกลุ่มให้แต่ละความหมายโดยใช้โปรแกรมหาค่าผลลัพธ์ที่ได้ค่าความถูกต้องมากที่สุด โดยส่วนที่แรงงในตารางคือกลุ่มที่ระบบเลือกให้เป็นตัวแทนของความหมายนั้น เนื่องจากเมื่อรวมค่าโดยรวมแล้วมีค่าความถูกต้องสูงสุด ซึ่งตารางที่ 5.5-5.8 นี้ ผู้วิจัยจะแสดงรายละเอียดเฉพาะตำแหน่งคำบริบทและกรอบหน้าต่างที่

ช่วยแยกความหมายได้ดีที่สุดในแต่ละคำเท่านั้น โดยรายละเอียดตำแหน่งคำบริบทและกรอบหน้าต่างอื่นนอกเหนือจากนี้สามารถดูได้จากภาคผนวก

ตารางที่ 5.5 รายละเอียดการจัดกลุ่มความหมายของ *เสียง* ด้วยคำบริบททางซ้าย 5 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. สิ่งที่ได้รับรู้ได้ด้วยหู	91	1736	47	1874
2. ความเห็น		51	2	53
3. คະแนนเสียง		89	2	91
รวม				2018

ตารางที่ 5.6 รายละเอียดการจัดกลุ่มความหมายของ *บอก* ด้วยคำบริบททางซ้าย 2 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	1421	80	138	1639
2. บ่งให้รู้	133	7	28	168
3. แนะนำ, สอน	184	4	23	211
รวม				2018

ตารางที่ 5.7 รายละเอียดการจัดกลุ่มความหมายของ *หัว* ด้วยคำบริบททางซ้าย 2 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. ส่วนบนสุดของ ร่างกายของคนหรือสัตว์	884	27		14	6	16	14	30	22	1013
2. ส่วนของพืชพันธุ์ บางอย่างตอนที่อยู่ใต้ ดิน	64					2	1	2	1	70
3. ส่วนแห่งสิ่งของ บางอย่างที่อยู่ข้างหน้า หรือข้างต้น	230			8		2	4	9	1	254

4. ส่วนแห่งสิ่งของที่ เป็นยอด	141			6		1		9	1	158
5. ผู้ที่มีความคิดหนักไป ทางใดทางหนึ่ง	54							2		56
6. ปัญญา, ความคิด	165			18		1	1	4	2	191
7. ผม	86		1						2	89
8. นามนัยแทนสิ่งที่ กล่าวถึงทั้งหมด	68		38							106
9. ส่วนของอุปกรณ์	61					3		5		69
รวม										2006

ตารางที่ 5.8 รายละเอียดการจัดกลุ่มความหมายของ *ติด* ด้วยคำบริบททางขวา 3 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. อาการที่ขี้อายอยู่		71	5	1		49		84	3	213
2. ทำให้เกิดขัดอยู่หรือ ขี้อายจนไม่อาจให้ เคลื่อนไปได้	11	84	6	1		7		189	5	303
3. ชอบอย่างขาดไม่ได้	4	2	7		1		49	114	23	200
4. ผนีก		21	5	34		13		153	1	227
5. ต่อเนื่องอย่างใกล้ชิด		15	150	1		140		127	2	435
6. ได้รับเชื้อโรค	1		4		273		1	29		308
7. คงอยู่หรือทำให้คง อยู่	7	65	3	9		28		88	1	201
8. อยู่ในที่คุมขัง	2	8	1					46	2	59
9. ผ่านเกณฑ์		5	1					54		60
รวม										2006

จากตารางที่ 5.5-5.8 เมื่อพิจารณาแล้ว สิ่งที่น่าสนใจคือระบบใช้อะไรในการแยกกลุ่มแต่ละกลุ่มออกจากกัน ซึ่งเมื่อดูตัวอย่างข้อมูลที่ระบบได้แยกออกเป็นแต่ละกลุ่มแล้วพบว่า ระบบใช้คำที่มีความถี่สูงในการปรากฏร่วมกับคำเป้าหมายเป็นตัวแบ่งกลุ่ม เช่น คำว่า *บอก* เมื่อใช้บริบททางซ้าย 2 ตำแหน่ง ระบบแบ่งข้อมูลออกเป็น 3 กลุ่ม โดยกลุ่มที่ 1 เป็นกลุ่มที่มีรูปแบบคำปรากฏร่วมไม่ชัดเจน กล่าวคือเป็นกลุ่มที่รวมคำปรากฏร่วมที่มีความถี่ในการปรากฏร่วมกับคำเป้าหมายน้อย กลุ่มที่ 2

ระบบใช้คำว่า *ไม่* เป็นตัวแบ่งกลุ่ม กล่าวคือ ในคำบริบททางซ้าย 2 ตำแหน่ง หากมีคำใดคำหนึ่งเป็นคำว่า *ไม่* ระบบจะจัดให้ตัวอย่างข้อมูลนั้นอยู่ในกลุ่มที่ 2 และในกลุ่มที่ 3 ระบบจัดกลุ่มโดยใช้คำว่า *ที่* กับ *จะ* ซึ่งหากมีคำใดคำหนึ่งจาก 2 คำเป็นคำว่า *ที่* กับ *จะ* ระบบก็จะจัดให้ตัวอย่างข้อมูลอยู่ในกลุ่มนี้เป็นต้น สาเหตุที่ระบบจัดกลุ่มเช่นนี้เนื่องจากเมื่อดูความถี่ของคำปรากฏร่วมกับคำว่า *บอก* ในบริบททางซ้ายแล้วจะพบว่า *ที่* จะ *ไม่* เป็นคำที่มีความถี่อยู่ในลำดับต้น ๆ ของรายการคำดังจะเห็นได้จากตารางที่ 5.9

ตารางที่ 5.9 ความถี่ของคำบริบททางซ้าย 2 ตำแหน่ง 5 คำแรกของ *บอก*

คำบริบททางซ้าย					บอก
ลำดับ	ตำแหน่งที่ 2		ตำแหน่งที่ 1		
1	ที่	67	จะ	149	
2	ไม่	60	ก็	129	
3	ก็	51	เขา	116	
4	แต่	34	มา	75	
5	เขา	30	ที่	51	

จากตารางจะเห็นว่าคำปรากฏร่วมที่อยู่ติดกับ *บอก* ในตำแหน่งที่ 1 ที่มีความถี่สูงที่สุดคือ *จะ* รองลงมาคือ *ก็* และในตำแหน่งที่ 2 คำที่มีความถี่สูงสุดคือ *ที่* รองลงมา คือ *ไม่* อย่างไรก็ตาม จะเห็นว่าในตำแหน่งที่ 1 คำว่า *ก็* และ *เขา* ต่างก็มีความถี่สูงเช่นกัน โดยมีความถี่สูงกว่า *ที่* และ *ไม่* ในตำแหน่งที่ 2 แต่ระบบไม่ได้นำเอาค่า 2 คำนี้มาใช้จัดกลุ่มด้วย เหตุที่เป็นเช่นนี้เพราะว่า *ก็* และ *เขา* เป็นคำที่ไปปรากฏร่วมกับคำอื่นมากมาย และมีความสัมพันธ์กับคำเหล่านั้นน้อย เช่น คำปรากฏร่วม *ก็-จะ* ปรากฏร่วมกัน 15 ครั้ง มีค่า log-likelihood เท่ากับ 5.86 ในขณะที่ *พอ-จะ* หรือ *อาจ-จะ* เกิดร่วมกันเพียง 9 และ 7 ครั้งตามลำดับซึ่งน้อยกว่า *ก็-จะ* แต่เมื่อหาค่า log-likelihood แล้วได้ค่าเท่ากับ 34.41 และ 22.26 ตามลำดับ จะเห็นว่าค่าสูงกว่าค่า log-likelihood ของ *ก็-จะ* ค่อนข้างมากนั้นแสดงว่า *พอ-จะ* และ *อาจ-จะ* มีความสัมพันธ์กันสูง ซึ่งระบบได้จัดให้ประโยคที่มีคำบริบทเป็น *อาจ-จะ* และ *พอ-จะ* อยู่ในกลุ่มที่ 3 ด้วย ในกรณีของคำอื่น ๆ ระบบก็จัดกลุ่มของตัวอย่างข้อมูลในลักษณะเดียวกัน คือ ใช้ความสัมพันธ์ระหว่างคำที่มีความสัมพันธ์สูงในการแยกกลุ่มความหมายเป็นหลัก เนื่องจากในการประมวลผลข้อมูลผู้วิจัยได้ใช้ข้อมูลที่ผ่านการแปลงค่าความถี่เป็นค่า log-likelihood แล้วไปใช้ เช่น คำว่า *หัว* ในตารางที่ 5.7 จะเห็นว่ากลุ่มที่ 2 ที่ระบบจัดกลุ่มไว้มีทั้งหมด 27 ตัวอย่าง เมื่อได้ดูตัวอย่างข้อมูลทั้งหมดที่ระบบได้แบ่งออกมาเป็นกลุ่มนี้ พบว่า ระบบใช้คู่คำปรากฏร่วม *เปลือก-คลุม* เป็นตัวแบ่งข้อมูล หากดูจากความถี่ของคำบริบททางซ้าย 2 ตำแหน่งตามทีระบบได้นำไปประมวลผลข้อมูล พบว่า *เปลือก* และ *คลุม* ไม่ใช่คำที่มีความถี่อยู่ในลำดับต้น ๆ ลำดับ

คำเรียงตามความถี่ของการปรากฏร่วมกับ หัว ในบริบททางซ้าย 2 ตำแหน่ง 10 คำแรก ได้แสดงไว้ในตารางที่ 5.10

ตารางที่ 5.10 ความถี่ของคำบริบททางซ้าย 2 ตำแหน่ง 10 คำแรกของ หัว

คำบริบททางซ้าย					หัว
ลำดับ	ตำแหน่งที่ 2		ตำแหน่งที่ 1		
1	มา	66	ใน	76	
2	ไป	48	ที่	68	
3	อยู่	45	ต่อ	67	
4	จะ	44	บน	58	
5	ที่	35	จาก	38	
6	ก็	28	มี	35	
7	เป็น	27	เอา	33	
8	เปลือก	25	ส่วน	32	
9	การ	24	คลุม	32	
10	มี	20	ลูบ	31	

จากตารางที่ 5.10 จะเห็นว่า ในตำแหน่งที่ 1 *คลุม* มีความถี่เป็นลำดับที่ 9 และ *เปลือก* ในตำแหน่งที่ 2 มีความถี่เป็นลำดับที่ 8 แต่คำ 2 คำนี้ปรากฏร่วมกันมากที่สุด โดยปรากฏร่วมกันทั้งหมด 25 ครั้ง และค่า log-likelihood มีค่าเท่ากับ 250.20 แสดงให้เห็นว่า *เปลือก* และ *คลุม* มีความสัมพันธ์กันสูงมาก ดังนั้นระบบจึงใช้ *เปลือก-คลุม* ในการแยกกลุ่มตัวอย่างข้อมูลออกมาเป็นอีก 1 กลุ่ม

อย่างไรก็ตาม จากภาพกราฟที่ 5.1-5.4 จะเห็นว่าส่วนใหญ่แล้วค่าความถูกต้องจะมีค่าลดลงเมื่อเพิ่มกรอบหน้าต่างมากขึ้น สาเหตุน่าจะมาจากเมื่อใช้กรอบหน้าต่างมากขึ้น คำปรากฏรวมก็จะมีหลากหลายมากขึ้นตามไปด้วย นั่นหมายความว่าคู่คำปรากฏรวมที่จะนำมาใช้ในการแบ่งกลุ่มก็จะมีมากขึ้นเช่นกัน ผู้วิจัยจะยกตัวอย่างการแบ่งกลุ่มของระบบด้วยบริบททางซ้าย 2 ตำแหน่ง 3 ตำแหน่ง และ 5 ตำแหน่งของคำว่า *บอก* ให้เห็นถึงการกระจายตัวของข้อมูล ดังแสดงในตารางที่ 5.11-5.13 ในตาราง ส่วนที่แรงงาคือคำตอบที่โปรแกรมเลือกให้เป็นตัวแทนของความหมายนั้น ๆ

ตารางที่ 5.11 รายละเอียดการจัดกลุ่มความหมายของ *บอก* ด้วยคำบริบททางซ้าย 2 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	1421	80	138	1639
2. บ่งให้รู้	133	7	28	168
3. แนะนำ, สอน	184	4	23	211
รวม	1738	91	189	2018

ตารางที่ 5.12 รายละเอียดการจัดกลุ่มความหมายของ *บอก* ด้วยคำบริบททางซ้าย 3 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	104	301	1234	1639
2. บ่งให้รู้	7	35	126	168
3. แนะนำ, สอน	11	47	153	211
รวม	122	383	1513	2018

ตารางที่ 5.13 รายละเอียดการจัดกลุ่มความหมายของ *บอก* ด้วยคำบริบททางซ้าย 5 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	738	532	369	1639
2. บ่งให้รู้	86	56	26	168
3. แนะนำ, สอน	92	64	55	211
รวม	916	652	450	2018

จากตารางที่ 5.11-5.13 จะเห็นว่า เมื่อใช้กรอบหน้าต่างมากขึ้น กลุ่มข้อมูลจะมีการกระจายไปตามกลุ่มต่าง ๆ มากขึ้น ดังจะเห็นได้จากจำนวนรวมของแต่ละกลุ่มที่ระบบแยกกลุ่มความหมายเมื่อใช้ตำแหน่งคำมากขึ้น ความต่างของจำนวนตัวอย่างในแต่ละกลุ่มลดลง เช่น เมื่อใช้คำบริบท 2 ตำแหน่ง กลุ่มที่ 1 จะมีจำนวนตัวอย่าง 1,738 ตัวอย่าง กลุ่มที่ 2 มี 91 ตัวอย่าง และกลุ่มที่ 3 มี 189 ตัวอย่าง ในขณะที่เมื่อใช้คำบริบท 5 ตำแหน่ง กลุ่มที่ 1 ถึง 3 มีจำนวนตัวอย่างเป็น 916 652 และ

450 ตามลำดับ เนื่องจากเมื่อใช้ตำแหน่งคำมากขึ้น คู่คำปรากฏรวมก็จะมีมากขึ้น ในขณะที่เดียวกันคู่คำที่มีความสัมพันธ์กันก็จะมีมากขึ้นตามไปด้วย ดังนั้นในแต่ละกลุ่มจึงมีคู่คำที่ใช้ในการแบ่งกลุ่มหลายคู่คำ จึงส่งผลทำให้ข้อมูลกระจายไปตามกลุ่มต่าง ๆ

5.2 จำนวนความหมายของคำ

เมื่อเปรียบเทียบค่าความถูกต้องระหว่างคำที่มีจำนวนความหมายน้อยกับคำที่มีจำนวนความหมายมาก พบว่าระบบสามารถแยกนัยความหมายของคำที่มีความหมายน้อยได้ถูกต้องมากกว่าคำที่มีนัยความหมายมาก ดังแสดงให้เห็นในตารางที่ 5.14 ดังนี้

ตารางที่ 5.14 ค่าความถูกต้องของระบบเมื่อแยกนัยความหมายของคำที่มีนัยความหมายน้อยกับคำที่มีนัยความหมายมาก

คำ บริบท	2 ตำแหน่ง				3 ตำแหน่ง				5 ตำแหน่ง			
	คำนาม		คำกริยา		คำนาม		คำกริยา		คำนาม		คำกริยา	
	น้อย	มาก	น้อย	มาก	น้อย	มาก	น้อย	มาก	น้อย	มาก	น้อย	มาก
	เสียง	หัว	บอก	ดีด	เสียง	หัว	บอก	ดีด	เสียง	หัว	บอก	ดีด
ซ้าย	1701	958	1453	479	1714	867	1288	518	1738	749	849	487
ขวา	1690	907	922	679	1494	785	973	811	1494	631	1201	739
ซ้าย-ขวา	1120	860	1001	617	1126	676	1089	688	978	531	849	624

หมายเหตุ: จำนวนข้อมูลทั้งหมดของ เสียง กับ บอก เท่ากับ 2018 ตัวอย่าง และ หัว กับ ดีด เท่ากับ 2006 ตัวอย่าง

จากตารางจะเห็นว่า ค่าความถูกต้องของระบบที่แยกคำที่มีนัยความหมายน้อยสูงกว่าระบบที่แยกคำที่มีนัยความหมายมากทั้งคำนามและคำกริยาในทุกตำแหน่งของคำบริบท ทั้งนี้เนื่องจากเมื่อคำมีนัยความหมายมาก นั้นหมายความว่าระบบจะต้องแบ่งกลุ่มข้อมูลออกเป็นหลายกลุ่ม ทำให้ข้อมูลกระจาย ดังนั้นเมื่อระบบที่แยกกลุ่มความหมายของคำที่มีนัยความหมายน้อยไม่ต้องกระจายข้อมูลออกเป็นหลายกลุ่มมากเกินไป ค่าความถูกต้องจึงมีสูงตามไปด้วย

เมื่อดูรายละเอียดจำนวนกลุ่มตัวอย่างที่ใช้ของแต่ละความหมายของแต่ละคำแล้ว ในคำที่มีนัยความหมายน้อย คือ เสียง และ บอก พบว่าทั้งสองคำจะมีความหมายหลักที่มีจำนวนตัวอย่างข้อมูลมากกว่าความหมายอื่นมาก ดังแสดงไว้ในตารางที่ 5.15 และจำนวนตัวอย่างของคำที่มีนัยความหมายมาก คือ หัว และ ดีด ได้แสดงไว้ในตารางที่ 5.16 ดังนี้

ตารางที่ 5.15 เปรียบเทียบการกระจายของจำนวนข้อมูลในแต่ละความหมายของ *เสียง* และ *บอก*

เสียง		บอก	
ความหมาย	ข้อมูล	ความหมาย	ข้อมูล
1. สิ่งที่เราได้ด้วยหู	1874	1. พูดให้รู้, เล่าให้ฟัง	1639
2. ความเห็น	53	2. บ่งให้รู้	168
3. คะแนนเสียง	91	3. แนะนำ, สอน	211
รวม	2018	รวม	2018

ตารางที่ 5.16 เปรียบเทียบการกระจายของจำนวนข้อมูลในแต่ละความหมายของ *หัว* และ *ติด*

หัว		ติด	
ความหมาย	ข้อมูล	ความหมาย	ข้อมูล
1. ส่วนบนสุดของร่างกายของคนหรือสัตว์	1013	1. อาการที่ช็อกอยู่	213
2. ส่วนของพืชพันธุ์บางอย่างตอนที่อยู่ใต้ดิน	70	2. ทำให้เกิดขัดอยู่หรือช็อกอยู่จนไม่อาจให้เคลื่อนไปได้	303
3. ส่วนแห่งสิ่งของบางอย่างที่อยู่ข้างหน้า หรือข้างต้น หรือแรกเริ่ม	254	3. ขอบอย่างขาดไม่ได้	200
4. ส่วนแห่งสิ่งของที่เป็นยอด ส่วนแห่งสิ่งของที่ยื่นเด่นออกไป	158	4. ผนึ้ก	227
5. ผู้ที่มีความคิดหนักไปทางใดทางหนึ่ง	56	5. ต่อเนื่องอย่างใกล้ชิด	435
6. ปัญญา, ความคิด	191	6. ได้รับเชื้อโรค	308
7. ผม	89	7. คงอยู่หรือทำให้คงอยู่	201
8. นามนัยแทนสิ่งที่กล่าวถึงทั้งหมด	106	8. อยู่ในที่คุมขัง	59
9. ส่วนของอุปกรณ์หรือเครื่องมือ	69	9. ผ่านเกณฑ์	60
รวม	2006	รวม	2006

จากตารางที่ 5.15 จะเห็นว่า ความหมายที่ 1 ของ *เสียง* และ *บอก* มีจำนวนตัวอย่างข้อมูลมากประมาณ 2 ใน 3 ของจำนวนข้อมูลทั้งหมด และความหมายที่ 1 ของ *หัว* ในตารางที่ 5.16 มีจำนวนตัวอย่างข้อมูลถึงประมาณครึ่งหนึ่งของข้อมูลทั้งหมด และแต่ละความหมายก็มีจำนวนตัวอย่างข้อมูลไม่เท่ากัน ดังนั้นข้อมูลที่มีจำนวนตัวอย่างน้อยเกินไปก็อาจทำให้ระบบไม่เห็นความสัมพันธ์ของคำปรากฏรวมได้ เช่น ความหมายที่ 8 ของคำว่า *ติด* ในตารางที่ 5.16 เมื่อดูจากข้อมูลตัวอย่าง พบว่า

ค่าปรากฏรวมที่มักพบในความหมายนี้ คือ *คุก ตะรวาง* แต่ระบบไม่สามารถเห็นความสัมพันธ์ได้ เนื่องจากจำนวนตัวอย่างมีน้อย เพียง 59 ตัวอย่างเท่านั้น แม้จะมีคู่ค่าปรากฏรวมที่มีความสัมพันธ์กันสูงแต่เนื่องจากมีจำนวนตัวอย่างน้อย ทำให้ค่า log-likelihood ที่ได้ไม่สูงพอที่จะทำให้ระบบเห็นความสัมพันธ์ได้ ดังนั้นระบบจึงไม่สามารถแยกความหมายนี้ออกมาได้

อย่างไรก็ตาม สำหรับความหมายที่มีจำนวนตัวอย่างข้อมูลมากดังเช่นความหมายที่ 1 ของ *เสียง บอก และ หัว* ดังแสดงในตารางที่ 5.15 และ 5.16 แม้ผลที่ออกมาพบว่าระบบสามารถแยกกลุ่มความหมายออกมาได้มากกว่าความหมายอื่นจริง ดังเช่นคำว่า *เสียง* ที่แสดงในตารางที่ 5.17 เป็นการจัดกลุ่มทางความหมายของระบบเมื่อใช้บริบททางซ้าย 2 ตำแหน่ง ส่วนที่แรเงาในตารางคือส่วนที่โปรแกรมเลือกให้เป็นคำตอบของความหมายนั้น ๆ

ตารางที่ 5.17 รายละเอียดการจัดกลุ่มความหมายของ *เสียง* ด้วยคำบริบททางซ้าย 2 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. สิ่งที่ได้รับรู้ได้ด้วยหู	1698	89	87	1874
2. ความเห็น	51	1	1	53
3. คະแนนเสียง	89		2	91
รวม				2018

จากตาราง จะเห็นว่าความหมายที่ 1 มีค่าความถูกต้องสูงที่สุด จากตัวอย่างทั้งหมด 1,874 ตัวอย่าง ระบบสามารถแยกกลุ่มออกมาได้ถึง 1,698 ตัวอย่าง แต่เมื่อได้วิเคราะห์กลุ่มตัวอย่างข้อมูลที่อยู่ในกลุ่มนี้แล้วพบว่าคำบริบทที่มีค่าความสัมพันธ์ไม่สูงจะถูกจัดให้อยู่ในกลุ่มนี้ ในขณะที่กลุ่มที่ 2 และ 3 ระบบจะแยกด้วยคู่ค่าที่มีค่า log-likelihood สูง ๆ เช่น ในกลุ่มที่ 2 ระบบแยกกลุ่มตัวอย่างด้วยคู่ค่า *ก็-ไต้ยีน* หรือ *เมื่อ-ไต้ยีน* ซึ่ง *ก็-ไต้ยีน* มีค่า log-likelihood เท่ากับ 96.75 และ *เมื่อ-ไต้ยีน* มีค่าเท่ากับ 57.84 เป็นต้น ดังนั้นจึงจะเห็นว่าระบบใช้ค่าความสัมพันธ์ระหว่างคำเป็นหลักในการแบ่งกลุ่มข้อมูล และเมื่อวิเคราะห์ตามคอลัมน์ในตารางที่ 5.17 จะเห็นว่าตัวอย่างข้อมูลเกือบทั้งหมดของทั้ง 3 ความหมายจะอยู่ในกลุ่มที่ 1 นั่นเป็นการแสดงให้เห็นว่าระบบไม่สามารถแยกความหมาย 3 ความหมายนี้ออกจากกันได้ ส่วนหนึ่งอาจเป็นเพราะจำนวนตัวอย่างของความหมายที่ 2 และ 3 มีน้อยมากเมื่อเทียบกับความหมายที่ 1 ดังนั้นระบบจึงไม่เห็นค่าปรากฏรวมในความหมายที่ 2 และ 3 เนื่องจากตัวอย่างมีไม่เพียงพอ

5.3 ประเภทของคำ

ในงานวิจัยนี้ได้ทดสอบให้ระบบแยกนัยความหมายของคำ 2 ประเภท คือ คำนาม กับ คำกริยา เพื่อเปรียบเทียบประสิทธิภาพของระบบว่าระบบสามารถแยกนัยความหมายของคำประเภทใดได้ดีกว่ากัน โดยเทียบคู่ของคำนามกับคำกริยาที่มีจำนวนความหมายใกล้เคียงกัน คือ *เสียง* คู่กับ *บอก* และ *หัว* คู่กับ *ติด* ดังแสดงในตารางที่ 5.18

ตารางที่ 5.18 ค่าความถูกต้องของระบบเมื่อแยกนัยความหมายของคำระหว่างคำนามและคำกริยา

คำ บริบท	2 ตำแหน่ง				3 ตำแหน่ง				5 ตำแหน่ง			
	ความหมายน้อย		ความหมายมาก		ความหมายน้อย		ความหมายมาก		ความหมายน้อย		ความหมายมาก	
	นาม	กริยา	นาม	กริยา	นาม	กริยา	นาม	กริยา	นาม	กริยา	นาม	กริยา
	เสียง	บอก	หัว	ติด	เสียง	บอก	หัว	ติด	เสียง	บอก	หัว	ติด
ซ้าย	1701	1453	958	479	1714	1288	867	518	1738	849	749	487
ขวา	1690	922	907	679	1494	973	785	811	1494	1201	631	739
ซ้าย-ขวา	1120	1001	860	617	1126	1089	676	688	978	849	531	624

หมายเหตุ: จำนวนข้อมูลทั้งหมดของ *เสียง* กับ *บอก* เท่ากับ 2018 ตัวอย่าง และ *หัว* กับ *ติด* เท่ากับ 2006 ตัวอย่าง

จะเห็นว่าส่วนใหญ่ค่าความถูกต้องของระบบที่แยกนัยความหมายของคำนามจะสูงกว่าระบบที่แยกนัยความหมายของคำกริยา โดยระบบที่ใช้คำบริบททางซ้ายในการแยกกลุ่มความหมายของคำนามจะสูงกว่าของคำกริยาในทุกตำแหน่ง จะมีเพียงบางบริบททางขวาและบริบททางซ้ายและขวาที่ค่าความถูกต้องของระบบที่แยกกลุ่มความหมายของคำกริยาจะสูงกว่าระบบของคำนาม

อย่างไรก็ตาม ผลที่ได้ไม่สามารถสรุปได้ว่าบริบทสามารถแยกความหมายของคำนามได้ดีกว่าคำกริยา เพราะจากการวิเคราะห์วิธีการแยกกลุ่มความหมายของระบบ พบว่าระบบใช้คำปรากฏร่วมที่มีค่าความสัมพันธ์กันสูงเป็นตัวแบ่งกลุ่ม ดังนั้นประเภทของคำจึงไม่มีผลต่อประสิทธิภาพของระบบในงานวิจัยนี้ หากว่าความหมายของคำมีคำปรากฏร่วมที่มีความสัมพันธ์กันมากก็จะทำให้ระบบสามารถแบ่งกลุ่มของความหมายได้ดี ดังตัวอย่างเปรียบเทียบคำว่า *หัว* และ *ติด* ในบริบททางขวา 3 ตำแหน่งที่ค่าความถูกต้องของ *ติด* มากกว่า *หัว* แม้จำนวนตัวอย่างข้อมูลของ *ติด* จะไม่มีความหมายใดเลยที่มีจำนวนตัวอย่างข้อมูลมากกว่าความหมายอื่นเป็นพิเศษเมื่อเทียบกับ *หัว* ที่ความหมายที่ 1 มีจำนวนตัวอย่างข้อมูลมากเป็นครึ่งหนึ่งของข้อมูลทั้งหมด ตารางที่ 5.19-5.20 แสดงรายละเอียดการจัดกลุ่มความหมายของคำว่า *หัว* และ *ติด* โดยใช้คำบริบททาง 3 ตำแหน่ง ส่วนที่แรเงาคือคำตอบที่โปรแกรมเลือกให้เป็นคำตอบของความหมายนั้น ๆ

ตารางที่ 5.19 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางขวา 3 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. ส่วนบนสุดของ ร่างกายของคนหรือสัตว์	37	22	49	715	7	55	25	39	64	1013
2. ส่วนของพืชพันธุ์ บางอย่างตอนที่อยู่ใต้ ดิน	4	3	3	51		5		3	1	70
3. ส่วนแห่งสิ่งของ บางอย่างที่อยู่ข้างหน้า หรือข้างต้น	13	11	8	176	1	22	4	9	10	254
4. ส่วนแห่งสิ่งของที่ เป็นยอด	5	3	6	126		11		5	2	158
5. ผู้ที่มีความคิดหนักไป ทางใดทางหนึ่ง	5	2	1	44		3		1		56
6. ปัญญา, ความคิด	5	12	23	138		5		3	5	191
7. ผม	6	6	4	71		2				89
8. นามนัยแทนสิ่งที่ กล่าวถึงทั้งหมด	12	2	15	67		3		6	1	106
9. ส่วนของอุปกรณ์	2	2	1	55		4		3	2	69
รวม										2006

ตารางที่ 5.20 รายละเอียดการจัดกลุ่มความหมายของ ติด ด้วยคำบริบททางขวา 3 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. อาการที่ข้องอยู่		71	5	1		49		84	3	213
2. ทำให้เกิดขัดอยู่หรือ ข้องอยู่จนไม่อาจให้ เคลื่อนไปได้	11	84	6	1		7		189	5	303
3. ขอบอย่างขาดไม่ได้	4	2	7		1		49	114	23	200
4. ผนึก		21	5	34		13		153	1	227
5. ต่อเนื่องอย่างใกล้ชิด		15	150	1		140		127	2	435

6. ได้รับเชื้อโรค	1		4		273		1	29		308
7. คงอยู่หรือทำให้คงอยู่	7	65	3	9		28		88	1	201
8. อยู่ในที่คุมขัง	2	8	1					46	2	59
9. ผ่านเกณฑ์		5	1					54		60
รวม										2006

จากตารางที่ 5.19 จะเห็นว่าการแบ่งกลุ่มข้อมูลของเครื่องจะค่อนข้างกระจาย ดังจะเห็นได้จากความหมายที่ 1 ที่ตัวอย่างข้อมูลถูกแบ่งกระจายออกไปทั้ง 9 กลุ่ม และตัวอย่างข้อมูลของความหมายส่วนใหญ่จะถูกจัดไว้ให้อยู่ในกลุ่มที่ 4 ซึ่งค่าปรากฏรวมที่จัดให้ตัวอย่างข้อมูลอยู่ในกลุ่มนี้คือ ค่าปรากฏรวมที่มีความสัมพันธ์กันน้อย นั้นแสดงให้เห็นว่าแม้จำนวนข้อมูลจะมาก แต่หากค่าที่ปรากฏร่วมกันมีความหลากหลายมากเกินไป ก็จะไม่มีความเด่นจนหาความสัมพันธ์ไม่ได้ เช่นความหมายที่ 1 มีจำนวนตัวอย่างทั้งหมด 1,013 ตัวอย่าง และถูกจัดให้มาอยู่ในกลุ่มที่ 4 ถึง 715 ตัวอย่าง หรือความหมายที่ 3 มี 254 ตัวอย่างแต่มาอยู่ในกลุ่มที่ 4 ถึง 176 ตัวอย่าง ซึ่งกลุ่มที่ 4 คือกลุ่มของข้อมูลที่ไม่มีความเด่นอะไรเป็นพิเศษ ในขณะที่คำว่า *ติด* ในตารางที่ 5.20 จะเห็นว่าไม่มีความหมายใดที่มีจำนวนมากกว่าความหมายอื่นจนเกินไป ดังจะเห็นได้จากจำนวนรวมในคอลัมน์สุดท้าย แม้ว่าจากตารางจะเห็นว่าตัวอย่างข้อมูลของทุกความหมายอยู่ในกลุ่มที่ 8 แต่ก็มีความหมายที่มีการแยกกลุ่มออกมาค่อนข้างชัดเจน เช่น กลุ่มที่ 5 ซึ่งเป็นคำตอบของความหมายที่ 6 ระบบได้ใช้ค่าปรากฏรวมที่มีคำว่า *เชื้อ* เป็นตัวแบ่งกลุ่ม หรือกลุ่มที่ 3 ซึ่งเป็นคำตอบของความหมายที่ 5 ระบบใช้ค่าปรากฏรวมที่มีคำว่า *กัน* และเครื่องไม้ไม้ยมก ๆ เป็นตัวแบ่งกลุ่ม เมื่อดูจำนวนตัวอย่างข้อมูลในกลุ่มที่ 5 แล้วพบว่าจำนวนตัวอย่างข้อมูลเกือบทั้งหมดมาจากความหมายที่ 6 หรือในกลุ่มที่ 3 แม้ว่าในกลุ่มนี้จะมีตัวอย่างข้อมูลจากความหมายอื่นจัดมาอยู่ในกลุ่มนี้ด้วย แต่ข้อมูลส่วนใหญ่ก็มาจากความหมายที่ 5 นั้นแสดงให้เห็นว่าหากตัวอย่างข้อมูลมีค่าปรากฏรวมที่ค่อนข้างชัดเจน มีความสัมพันธ์ระหว่างค่าสูง ระบบก็สามารถแยกกลุ่มความหมายได้โดยที่จำนวนตัวอย่างข้อมูลไม่จำเป็นต้องมีมากจนเกินไป เพราะแม้มีมากแต่หากเป็นค่าปรากฏรวมที่ไม่มีความเด่นหรือมีความหลากหลายมากเกินไป ก็จะทำให้ไม่เห็นความสัมพันธ์ของค่าปรากฏรวมที่จะช่วยแยกกลุ่มข้อมูล

5.4 บริบทที่เครื่องใช้แยกนัยความหมายของคำที่ใช้ในงานวิจัย

สำหรับการแยกกลุ่มตัวอย่างข้อมูลด้วยเครื่องในงานวิจัยนี้ สิ่งสำคัญคือค่าปรากฏรวมที่มีความสัมพันธ์กันสูง คือเป็นค่าที่ปรากฏร่วมกันบ่อย และมักไม่พบว่าเป็นค่าปรากฏร่วมกับค่าอื่น โดยค่าที่นำมาใช้แสดงความสัมพันธ์ของคำในงานวิจัยนี้คือ ค่า $\log\text{-likelihood}$ เมื่อเป็นเช่นนั้น ค่าปรากฏรวม

ใดที่มีความถี่และความสัมพันธ์กันสูง ระบบก็จะนำมาใช้ในการแยกกลุ่มความหมาย เช่น คำว่า *บอก* เมื่อให้ระบบแยกกลุ่มความหมายออกเป็น 3 กลุ่ม โดยใช้คำบริบททางขวา 2 ตำแหน่ง พบว่า ระบบใช้คำปรากฏร่วมที่มีคำว่า *ว่า* แยกตัวอย่างข้อมูลออกไป 1 กลุ่ม และใช้คำปรากฏร่วมที่มีคำว่า *ไม่* แบ่งข้อมูลออกไปเป็นกลุ่มที่ 2 และกลุ่มสุดท้ายคือกลุ่มที่รวมคำปรากฏร่วมอื่น ๆ ที่เหลือ เมื่อพิจารณาคำปรากฏร่วมที่ระบบนำมาใช้ในการแยกกลุ่มความหมายของ *บอก* จะเห็นว่าคำว่า *ว่า* หรือ คำว่า *ไม่* เป็นคำที่ไม่จำเพาะเจาะจงไปที่ความหมายใดความหมายหนึ่ง สามารถพบได้ในทุกความหมายของ *บอก* เช่น

ความหมายที่ 1 พุดให้รู้, เล่าให้ฟัง เช่น

|ณ|วัน|นี้|พ่อ|คำ|**บอก**|ว่า|กิน|น้ำ|ลูก|ยอ|แล้ว|รักษา|โรค|ได้|สาร|พัด|ชนิด|

|เพื่อ|ณ|ฝรั่ง|ที่|นั้น|**บอก**|ว่า|คน|ไทย|ที่|มา|เรียน|ระดับ|ปริญญา|โท|

ความหมายที่ 2 บ่งให้รู้ เช่น

|การ|วาง|นิ้ว|ชี้|ไว้|ที่|ริม|ฝี|ปาก|เพื่อ|**บอก**|ว่า|ให้|เสียบ|เป็น|ต้น|

|สัญญา|ตัญญาณ|ห่วย|ๆ|ของ|ฉัน|**บอก**|ชัดเจน|ว่า|เธอ|มี|ความ|สำคัญ|กับ|เขา|

ความหมายที่ 3 แนะนำ, สอน เช่น

|คุณ|หมอ|**บอก**|ว่า|อย่า|ไป|เล่น|แคม|ฟอก|มาก|

|คุณ|อัน|นาย|เก่า|ชอบ|**บอก**|ว่า|ให้|นึก|ว่า|ร้าน|ของ|เรา|เป็น|บ้าน|ของ|เรา|

ดังนั้นในความเป็นจริง ระบบไม่สามารถแยกความแตกต่างทางความหมาย 3 ความหมายของคำว่า *บอก* ได้ เนื่องจากไม่ใช่คำปรากฏร่วมที่มีความถี่และความสัมพันธ์กันสูงทุกคำจะสามารถแยกกลุ่มความหมายได้ เนื่องจากไม่ได้มีส่วนช่วยในการบ่งบอกความหมาย อย่างไรก็ตาม เนื่องจากในงานวิจัยนี้ใช้เพียงรูปของคำปรากฏร่วมเท่านั้นในการช่วยแยกกลุ่มความหมาย ดังนั้นเฉพาะความหมายที่มีคำที่ปรากฏร่วมชัดเจนและพบในความหมายอื่นน้อย รวมถึงมีจำนวนตัวอย่างข้อมูลมากพอที่จะทำให้เห็นความสัมพันธ์ระหว่างคำชัดเจน ระบบก็จะสามารถแยกความหมายนั้นได้ ดังเช่น คำว่า *เชื้อ* ที่จะพบในความหมายว่า “ได้รับเชื้อโรค” ของคำว่า *ติด* และแทบจะไม่พบในความหมายอื่นเลย ดังนั้นระบบก็จะสามารถแยกกลุ่มความหมายนี้ได้

สำหรับคำส่วนใหญ่ที่ไม่สามารถแยกความหมายได้มีทั้งในกรณีที่จำนวนตัวอย่างข้อมูลมาก แต่เนื่องจากคำมีความหลากหลาย มีการกระจายมากเกินไป จนไม่มีคำปรากฏร่วมใดโดดเด่น ระบบจึงไม่สามารถแยกกลุ่มความหมายได้ กับอีกกรณีหนึ่งคือจำนวนตัวอย่างข้อมูลมีน้อยเกินไป แม้ว่าคำ

ปรากฏว่านั่นจะมีความสัมพันธ์กันสูง แต่เนื่องจากข้อมูลมีไม่เพียงพอ จึงทำให้ไม่เห็นความเด่น ซึ่งในกรณีนี้ ระบบก็ไม่สามารถแยกกลุ่มความหมายได้เช่นกัน

ในบทนี้ ผู้วิจัยได้ทดลองแยกความหมายของคำด้วยเครื่อง ซึ่งเครื่องก็สามารถแยกความหมายของคำได้ในระดับหนึ่ง แต่เนื่องจากคุณสมบัติที่ผู้วิจัยใช้ในงานวิจัยนี้มีเพียงรูปคำของคำปรากฏเท่านั้น ดังนั้นประสิทธิภาพโดยรวมของระบบจึงยังไม่ดีเท่าที่ควร ผลการทดลองด้วยเครื่องนี้ผู้วิจัยจะนำไปเปรียบเทียบกับการวิเคราะห์ความหมายทางภาษาศาสตร์ที่ได้วิเคราะห์ไว้ในบทที่ 4 ในบทถัดไป

บทที่ 6

สรุป อภิปราย และข้อเสนอแนะ

ในส่วนนี้ผู้วิจัยจะสรุปผลการวิจัยทั้งหมด และอภิปรายผลการวิจัย รวมถึงนำเสนอข้อเสนอแนะที่จะเป็นประโยชน์ต่องานวิจัยต่อไป

6.1 สรุปผลการวิจัย

ในงานวิจัยนี้ ได้นำเสนอระบบการแยกนัยความหมายของคำหลายความหมาย โดยใช้แนวทางการวิเคราะห์ความหมายแบบแฝง โดยคำที่ศึกษามี 4 คำ แบ่งออกเป็นคำนาม 2 คำ ได้แก่ *เสียง* กับ *หัว* และคำกริยา 2 คำ ได้แก่ *บอก* กับ *ติด* โดยที่คำว่า *เสียง* และ *บอก* เป็นตัวแทนของคำที่มีนัยความหมายน้อย และ *หัว* และ *ติด* เป็นตัวแทนของคำที่มีนัยความหมายมาก โดยได้ตั้งสมมติฐานไว้ว่า ระบบจะสามารถแยกนัยความหมายของคำที่มีนัยความหมายน้อยได้ถูกต้องมากกว่าคำที่มีนัยความหมายมาก และสามารถแยกนัยความหมายของคำนามได้ถูกต้องมากกว่าคำกริยา รวมถึงบริบทที่เหมาะสมในการแยกนัยความหมายของคำนามคือบริบททางขวา และบริบทที่เหมาะสมในการแยกนัยความหมายของคำกริยา คือบริบททั้งทางซ้ายและทางขวา

ตัวอย่างข้อมูลที่ใช้มาจากคลังข้อมูลภาษาไทยแห่งชาติ โดยที่ *เสียง* และ *บอก* มีจำนวนตัวอย่างข้อมูลคำละ 2,018 ตัวอย่าง และ *หัว* และ *ติด* มีจำนวนตัวอย่างข้อมูลคำละ 2,006 ตัวอย่าง ก่อนทำการทดลองแยกความหมายด้วยเครื่อง ผู้วิจัยได้วิเคราะห์ความหมายของคำทั้ง 4 คำก่อน เพราะความหมายที่พบในพจนานุกรมอาจไม่ครอบคลุมความหมายทั้งหมด โดยการวิเคราะห์ความหมายผู้วิจัยจะดึงข้อมูลออกมาจากคลังข้อมูลคราวละ 200 ตัวอย่างและจะวิเคราะห์ความหมายจนกว่าจะไม่เจอความหมายใหม่จากข้อมูล 2 ครั้งติดกัน ความหมายที่วิเคราะห์ได้มีทั้งความหมายที่พบในพจนานุกรม และมีบางความหมายในพจนานุกรมที่ไม่พบในงานวิจัยนี้ รวมถึงมีความหมายใหม่ที่เพิ่มเติมขึ้นมานอกเหนือจากในพจนานุกรม ความหมายของคำที่ได้ หากความหมายนั้นมีจำนวนตัวอย่างไม่ถึง 50 ตัวอย่าง ผู้วิจัยจะไม่นำความหมายนั้นมาทดลองแยกความหมายด้วยเครื่อง สรุปจำนวนความหมายที่นำมาใช้ของ *เสียง* และ *บอก* เท่ากับ 3 ความหมาย และ *หัว* และ *ติด* เท่ากับ 9 ความหมาย

ในการทดลองแยกความหมายของคำด้วยเครื่อง คำบริบทที่ใช้ในการทดสอบคือ คำบริบททางซ้าย คำบริบททางขวา และคำบริบททั้งทางซ้ายและทางขวา โดยกำหนดกรอบหน้าต่างของคำบริบทเป็นช่วง 2 ตำแหน่ง 3 ตำแหน่ง และ 5 ตำแหน่งจากคำเป้าหมาย ในการประเมินค่าความถูกต้องผู้วิจัยให้โปรแกรมคำนวณการเรียงสับเปลี่ยนคำตอบเพื่อหาคำตอบแบบใดให้ผลดีที่สุด

จากผลการทดสอบพบว่า บริบททางด้านการช่วยแยกนัยความหมายได้ดีที่สุดในคำว่า *เสี่ยง* และ *หัว* ในขณะที่บริบททางขวาช่วยแยกนัยความหมายของคำว่า *ติด* ได้ดีกว่าบริบทอื่น และคำว่า *บอก* บริบททางซ้ายช่วง 2 ตำแหน่งและ 3 ตำแหน่ง จะช่วยแยกนัยความหมายได้ดีที่สุด แต่เมื่อเปลี่ยนเป็นช่วงบริบท 5 ตำแหน่ง บริบททางขวาจะช่วยแยกนัยความหมายได้ดีที่สุด สาเหตุที่ใช้ช่วงคำบริบทน้อยกว่าช่วงบริบทที่มากขึ้น เป็นเพราะว่า เมื่อยังใช้ช่วงคำบริบทที่มากขึ้น คำปรากฏร่วมก็จะมีหลากหลายมากขึ้น ทำให้เกิดการกระจายของข้อมูลสูง ส่งผลให้ระบบไม่เห็นความสัมพันธ์ระหว่างคำเป้าหมายกับคำปรากฏร่วมจึงทำให้ประสิทธิภาพของระบบลดลงเมื่อใช้กรอบหน้าต่างที่มากขึ้น เมื่อวิเคราะห์ถึงสิ่งที่ช่วยระบบในการแยกกลุ่มความหมาย พบว่า ระบบใช้รูปแบบของคำปรากฏร่วมที่มีความถี่และความสัมพันธ์ระหว่างคำสูงเป็นตัวช่วยแบ่งกลุ่มความหมาย ซึ่งความสัมพันธ์ระหว่างคำนี้มาจากค่า $\log\text{-likelihood}$ ที่ได้จากการแปลงค่าความถี่ก่อนนำข้อมูลไปประมวลผล เป็นค่าที่แสดงให้เห็นว่าคำปรากฏร่วมคู่ใดมีความสัมพันธ์กัน คำที่ปรากฏร่วมกันถี่และปรากฏร่วมกับคำอื่นน้อย ค่า $\log\text{-likelihood}$ จะสูงเพราะคำมีความสัมพันธ์กันมาก แต่ถ้าคำปรากฏร่วมกันถี่และปรากฏร่วมกับคำอื่นถี่เช่นกัน ค่า $\log\text{-likelihood}$ จะต่ำเพราะคำมีความสัมพันธ์กันน้อย เช่น คำว่า *ติด* ในความหมายว่า “ได้รับเชื้อโรค” มักปรากฏร่วมกับคำว่า *เชื้อ* เสมอเป็นคำว่า *ติดเชื้อ* โดยที่ *เชื้อ* จะปรากฏอยู่ในความหมายอื่นน้อยมาก ดังนั้นระบบจึงใช้คำว่า *เชื้อ* เป็นตัวแยกกลุ่มความหมายด้วย ดังจะเห็นได้จากตารางที่ 6.1 ซึ่งแสดงการจัดกลุ่มด้วยเครื่องโดยใช้คำบริบททางขวา 2 ตำแหน่ง

ตารางที่ 6.1 รายละเอียดการจัดกลุ่มความหมายของ *ติด* ด้วยคำบริบททางขวา 2 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. อาการที่ช็อกอยู่	140	32	1	18		4			18	213
2. ทำให้เกิดช็อคอยู่หรือ ช็อกอยู่จนไม่อาจให้ เคลื่อนไปได้	216	2	1	62		2	2		18	303
3. ชอบอย่างขาดไม่ได้	177		17	1	1	2	2			200
4. ผนึก	205	3		17		2				227
5. ต่อเนื่องอย่างใกล้ชิด	263	10		4		87	46	25		435
6. ได้รับเชื้อโรค	34		1	1	270	2				308
7. คงอยู่หรือทำให้คง อยู่	134	9	1	50		1			6	201
8. อยู่ในที่คุมขัง	52			6					1	59

9. ผ่านเกณฑ์	56			1					3	60
รวม	1277	56	21	160	271	100	50	25	46	2006

จากตารางจะเห็นว่าข้อมูลในกลุ่มที่ 5 มีจำนวนรวมเท่ากับ 271 โดยที่เป็นตัวอย่างข้อมูลจากความหมายที่ 6 ถึง 270 ตัวอย่าง เมื่อดูรายละเอียดคำปรากฏรวมของตัวอย่างข้อมูลที่อยู่ในกลุ่มนี้จะพบว่าเป็นคำที่ปรากฏร่วมกับ เชื้อ เช่น

|ติด|เชื้อ|แมลง|เจาะ|สัตว์|แพะ|หรือ|มนุษย์|

|ติด|เชื้อ|รวม|ทั้ง|โรค|ผิวหนัง|ที่|มี|อาการ|อักเสบ|

|ติด|เชื้อ|ไวรัส|ของ|ระบบ|ทาง|เดิน|อาหาร|

|ติด|เชื้อ|อย่าง|รุนแรง|ซึ่ง|แพร่|กระจาย|เข้าไป|

|ติด|เชื้อ|เอดส์|จริง|ๆ|หรือ|รัง|เกียจ|ที่|เธอ|เป็น|กะ|เทย|

6.2 อภิปรายผลการวิจัย

6.2.1 คำกริยามีนัยความหมายหลากหลายมากกว่าคำนามหรือไม่

จากการวิเคราะห์ความหมายของคำทั้ง 4 คำตามแนวทางภาษาศาสตร์พบว่าคำกริยามีนัยความหมายหลากหลายมากกว่าคำนาม ซึ่งจะเห็นได้ชัดเจนในคำที่มีนัยความหมายมาก โดยความหมายของคำว่า ติด ที่วิเคราะห์ได้จะมีอยู่ 18 ความหมาย และคำว่า หัว มี 11 ความหมาย ส่วนใหญ่ความหมายของ ติด ที่วิเคราะห์ได้จะผันไปตามกรรมในประโยค โครงสร้างประโยคของติดจะเป็นดังนี้

นามวลี (1) + ติด + นามวลี (2)

ตัวอย่าง คำว่าติดในความหมายว่า “ผ่านเกณฑ์” นามวลี (2) ที่จะปรากฏร่วมกับ ติด ในความหมายนี้ มักเป็นคำที่มีความหมายเกี่ยวกับองค์กร เช่น

|ทั้ง|สอง|คน|ก็|เอ็น|ที|ติด|มหาวิทยาลัย|เดียวกัน|ใน|กรุงเทพฯ|

|ก็มี|สิทธิ|เข้ามา|ติด|ทีม|ชาติ|ไทย|ชุด|ใหญ่|อย่าง|แน่นอน|

จะเห็นว่า จากตัวอย่าง ทั้ง มหาวิทยาลัย และ ทีมชาติไทย ต่างมีลักษณะเป็นองค์กร เพราะมีโครงสร้างการบริหารจัดการกันภายใน มีผู้บริหาร ผู้จัดการ เป็นต้น หรืออีกความหมายหนึ่งคือ “ติดตั้ง” นามวลี (2) ที่ปรากฏร่วมกับ ติดส่วนใหญ่จะเป็นคำที่หมายถึง พวกอุปกรณ์ เครื่องกลไกต่าง ๆ ที่ต้องอาศัยการประกอบจึงจะใช้งานได้ เช่น แอร์ เครื่องทำน้ำร้อน หัวเจาะ กล้อง เป็นต้น เช่น

|ค่าแรง|อะไร|มัน|แพง|ไป|หมด| |ร้าน|ก็|**ติด**|แอร์|ไม่|จิ้น|สาว| |ฯ| |ไม่|เข้า|

|ตอน|บน|ของ|ผนัง|ห้อง|ด้าน|หนึ่ง|**ติด**|ลูก|กรง|เหล็ก| |เว้น|ช่อง|ห่าง|พอ|

จะเห็นว่าความหมายของ **ติด** จะผันไปกับอาร์กิวเมนต์ที่ปรากฏร่วมหรือกรรมของประโยค ดังนั้นความหมายใหม่จึงมีโอกาสดำเนินมาได้ง่ายกว่าคำนาม เพราะความหมายของคำกริยามีลักษณะเป็นปลายเปิด(open-ended meaning) (Falkum and Vicente, 2015)

อย่างไรก็ตาม ในงานวิจัยนี้มีข้อจำกัดเรื่องการดึงข้อมูลออกมาวิเคราะห์ความหมาย คือ จะดึงข้อมูลออกมาคราวละ 200 ตัวอย่างแล้ววิเคราะห์ความหมายจนกว่าจะไม่พบความหมายใหม่ 2 ครั้งติดกัน ซึ่งหากผู้วิจัยพบความหมายใหม่หลังจากนี้ ก็จะไม่นำมารวมไว้เป็นความหมายของคำนั้น ๆ ดังนั้นจึงอาจเป็นสาเหตุหนึ่งที่ทำให้ความหมายของ **หั่ว** น้อยกว่า **ติด** เพราะเป็นไปได้ว่าในช่วงข้อมูลที่ดึงมาตอนแรกยังไม่พบความหมายที่หลากหลาย แต่ความหมายใหม่อาจจะมีเพิ่มขึ้นจากข้อมูลในช่วงหลังจากที่ผู้วิจัยไม่วิเคราะห์ความหมายใหม่เพิ่มแล้ว ซึ่งหากได้วิเคราะห์ความหมายจากข้อมูลทั้งหมดแล้วก็อาจเป็นไปได้ว่าความหมายของ **หั่ว** อาจจะเท่ากับ **ติด** หรือมากกว่า **ติด** ก็ได้

6.2.2 ระบบแยกคำที่มีนัยความหมายน้อยได้ดีกว่าคำที่มีนัยความหมายมากหรือไม่

เมื่อใช้เครื่องในการแยกความหมายของคำ พบว่า ระบบที่ใช้กับคำที่มีนัยความหมายน้อยจะมีความถูกต้องมากกว่าระบบที่ใช้กับคำที่มีนัยความหมายมาก เนื่องจากว่าเมื่อจำนวนความหมายมาก ระบบจะต้องกระจายข้อมูลออกเป็นหลายกลุ่ม ทำให้ค่าความถูกต้องลดลง เมื่อเทียบกับระบบที่กระจายข้อมูลน้อยกว่า ทั้งนี้เพราะจำนวนรวมของตัวอย่างของคำที่มีนัยความหมายน้อยกับคำที่มีนัยความหมายมากมีปริมาณใกล้เคียงกัน คือจำนวน 2,018 และ 2,006 ตามลำดับ นั่นหมายความว่าหากมีการแบ่งจำนวนกลุ่มตัวอย่างให้แต่ละความหมายเท่า ๆ กัน จำนวนตัวอย่างข้อมูลในแต่ละความหมายของคำที่มีนัยความหมายมากย่อมน้อยกว่าคำที่มีนัยความหมายน้อย ซึ่งจะส่งผลให้ระบบไม่เห็นรูปแบบของคำปรากฏร่วมที่จะสามารถช่วยแยกความหมายนั้น ๆ ออกจากความหมายอื่นเนื่องจากปริมาณตัวอย่างข้อมูลมีน้อยเกินไป

เมื่อพิจารณารายละเอียดของจำนวนตัวอย่างข้อมูลของความหมายของคำแต่ละคำแล้วพบว่าตัวอย่างข้อมูลส่วนใหญ่ของ **เสียด** และ **บอก** จะไปอยู่ที่ความหมายที่ 1 คือประมาณ 2 ใน 3 ของข้อมูลทั้งหมดนั้นหมายความว่าอีก 2 ความหมายจะมีจำนวนตัวอย่างข้อมูลน้อยมากเมื่อเทียบกับความหมายที่ 1 ส่วนคำที่มีนัยความหมายมาก ความหมายของ **หั่ว** จะไปอยู่ที่ความหมายที่ 1 คือประมาณครึ่งหนึ่งของข้อมูล แต่คำว่า **ติด** จะไม่มีความหมายใดความหมายหนึ่งที่มีจำนวนตัวอย่างมากจนเกินไป แม้ว่าเมื่อได้วิเคราะห์ความหมายตามแนวภาษาศาสตร์ของคำทั้ง 4 คำ จะทำให้เห็นว่ามีความหมายหลายความหมายที่จะมีคำปรากฏร่วมที่สามารถบ่งชี้ได้ว่าเป็นความหมายนั้น ๆ เช่น

ความหมายว่า “ความนิยม คะแนนเสียง” ของคำว่า *เสียง* จะพบว่า *เสียง* ในความหมายนี้มักปรากฏร่วมกับคำว่า *ข้างมาก* ในบริบททางขวา หรือคำว่า *หัว* ในความหมายว่า “ข้อความสำคัญของเรื่อง หรือข่าว” มักปรากฏร่วมกับคำกริยา *พาด จั่ว ไปรย* หรือความหมายว่า “อาการที่ซึ้งอยู่” ของคำว่า *ติด* มักปรากฏร่วมกับคำว่า *ยึด* ในบริบททางด้านซ้ายหรือขวาเป็นต้น แต่เนื่องจากความหมายเหล่านี้มีจำนวนตัวอย่างข้อมูลน้อยเกินไป เช่น ความหมายว่า “ความนิยม คะแนนเสียง” ของคำว่า *เสียง* มีจำนวนตัวอย่างเพียง 91 ตัวอย่าง ดังนั้นถึงแม้ว่าในความหมายนั้นจะมีค่าปรากฏร่วมที่จะพบได้ในความหมายนี้เท่านั้น แต่เมื่อความถี่ที่ปรากฏร่วมกันมีน้อย ระบบก็จะไม่เห็นถึงความสัมพันธ์และไม่สามารถแยกความหมายนี้ออกมาได้

กล่าวโดยสรุปคือ ระบบสามารถแยกกลุ่มความหมายของคำที่มีนัยความหมายน้อยได้ดีกว่าคำที่มีนัยความหมายมาก เนื่องจากการกระจายของข้อมูลในแต่ละกลุ่มน้อยกว่า

6.2.3 ระบบสามารถแยกนัยความหมายของคำนามได้ถูกต้องมากกว่าคำกริยาหรือไม่

หลังจากได้ทำการทดลองให้ระบบแยกกลุ่มความหมายแล้ว พบว่า ระบบใช้ความสัมพันธ์ของค่าปรากฏร่วมในการแยกกลุ่มความหมาย กล่าวคือ ถ้าค่าที่ปรากฏร่วมกันมีความถี่สูงและมีค่าความสัมพันธ์หรือค่า log-likelihood สูงระบบก็จะนำเอาค่าปรากฏร่วมนี้มาใช้ในการแยกกลุ่มความหมาย ดังนั้นหากความหมายใด ที่มีค่าปรากฏร่วมเกิดกับความหมายนี้ในความถี่ที่สูง และไม่ปรากฏหรือไปปรากฏร่วมกับความหมายอื่นน้อย ระบบก็จะนำค่าปรากฏร่วมนั้นมาใช้ในการแยกความหมาย ดังเช่น คำปรากฏร่วม *ๆ-กัน* ที่จะปรากฏร่วมกับคำว่า *ติด* เป็นกริยาลี *ติด ๆ กัน* ในความหมายว่า “ต่อเนื่องอย่างใกล้ชิด” ซึ่งไม่พบว่า *ๆ-กัน* ไปปรากฏในความหมายอื่น ดังนั้นระบบจึงสามารถแยกกลุ่มข้อมูลในความหมายนี้ออกมาได้

ในงานวิจัยนี้ จะเห็นว่าผลที่ได้ จะมีบางบริบทที่ประสิทธิภาพของระบบที่ใช้แยกความหมายของคำนามดีกว่า และขณะเดียวกันก็มีบางบริบทที่ค่าความถูกต้องของระบบที่ใช้แยกความหมายของคำกริยาสูงกว่าคำนาม ในความเป็นจริง เราไม่สามารถสรุปได้ว่าระบบสามารถแยกนัยความหมายของคำนามได้ถูกต้องมากกว่าคำกริยา หรือคำกริยาถูกต้องมากกว่าคำนาม เพราะแท้จริงแล้วประสิทธิภาพของระบบขึ้นอยู่กับคำที่เราเลือกนำมาใช้ในทดสอบระบบ เพราะคำแต่ละคำส่วนใหญ่มีจำนวนความหมายไม่เท่ากัน หรือถึงมีความหมายเท่ากันก็มีรายละเอียดต่อไปว่า แต่ละความหมายมีจำนวนตัวอย่างมากน้อยเพียงใด เพียงพอไหม มีค่าปรากฏร่วมที่ปรากฏร่วมด้วยถี่ไหม ซึ่งปัจจัยทุกอย่างที่กล่าวมานี้ล้วนส่งผลต่อประสิทธิภาพของระบบทั้งสิ้น ดังนั้นจึงสรุปได้ว่าประสิทธิภาพของระบบขึ้นอยู่กับคำที่เลือกมาทดสอบมากกว่า

6.2.4 บริบทที่เหมาะสมในการแยกนัยความหมายของคำนามและคำกริยา

จากผลการทดลองพบว่าบริบททางซ้ายช่วยแยกนัยความหมายได้ดีกว่าบริบทอื่นในคำว่า *เสียง* และ *หัว* ในขณะที่บริบททางขวาช่วยแยกนัยความหมายของคำว่า *ติด* ได้ดีกว่าบริบทอื่นและคำว่า *บอก* ขึ้นอยู่กับช่วงบริบทที่ใช้ คือ ช่วงบริบท 2 ตำแหน่งและ 3 ตำแหน่ง บริบททางซ้ายจะช่วยแยกนัยความหมายได้ดีที่สุด และช่วงบริบท 5 ตำแหน่ง บริบททางขวาจะช่วยแยกนัยความหมายได้ดีกว่าทางซ้าย

แม้ว่าผลการทดลองจะออกมาว่าคำนาม *เสียง* และ *หัว* บริบททางซ้ายจะช่วยแยกนัยความหมายได้ดีกว่าบริบททางขวา แต่นั่นไม่ได้หมายความว่าบริบททางซ้ายจะมีความเหมาะสมมากกว่าในการช่วยเครื่องแยกความหมายของคำนาม เนื่องจากในงานวิจัยนี้ ใช้เพียงแค่รูปคำ ไม่ได้มีคุณสมบัติทางด้านภาษาใด ๆ มาใช้ช่วยเครื่องในการแยกความหมาย ดังนั้นสิ่งที่เครื่องสามารถนำมาใช้ได้จึงมีเพียงความถี่ของคำที่ปรากฏร่วมกัน และความสัมพันธ์ระหว่างคำที่สามารถคำนวณได้จากความถี่ของคำ 2 คำที่ปรากฏร่วมกันเทียบกับเมื่อไปปรากฏร่วมกับคำอื่น ตัวอย่างคำว่า *เสียง* ที่เมื่อใช้บริบท 2 ตำแหน่งทางซ้ายในการแยกความหมาย เมื่อได้ดูคำที่เครื่องใช้ช่วยในการแยกกลุ่ม จะพบว่าไม่ได้เป็นคำที่สามารถบ่งชี้ว่าเป็นความหมายใดเป็นพิเศษ เช่น *ก็-ได้ยิน เมื่อ-ได้ยิน ไป-ตามตาม-ด้วย* เป็นต้น จากตัวอย่างที่ยกมาจะเห็นว่าคำปรากฏร่วมเหล่านี้เป็นคำทั่วไปที่สามารถพบได้กับทุกความหมายของ *เสียง* แต่เนื่องจากคำเหล่านี้มีความถี่ในการปรากฏสูง และอาจปรากฏร่วมกันมากกว่าคำอื่น จึงมีค่าความสัมพันธ์ระหว่างคำสูง เมื่อเป็นเช่นนี้ จึงไม่อาจสรุปได้ว่าบริบททางซ้ายมีความเหมาะสมในการแยกกลุ่มความหมายได้ดีกว่าบริบททางขวา เพราะเมื่อวิเคราะห์ความหมายตามหลักภาษาศาสตร์แล้วพบว่าในคำว่า *เสียง* และ *หัว* บริบททางขวาก็ยังคงช่วยบ่งชี้ความหมายได้ดีกว่าบริบททางซ้าย เช่น คำว่า *เสียง* ในความหมายว่า “สิ่งที่รับรู้ได้ด้วยหู” คำขยายที่อยู่ด้านขวา เช่น *ดัง หัวเราะ ร้อง เพลง กระซิบ* เป็นต้น สามารถเป็นสิ่งที่บ่งชี้ความหมายนี้ได้ แต่เนื่องจากในตัวอย่างข้อมูลที่นำมาใช้ในการทดสอบระบบ คำในความหมายนี้อาจมีการกระจายมากเกินไป ดังนั้นระบบจึงไม่เห็นความสัมพันธ์ของคำ หรือคำว่า *เสียง* ในความหมายว่า “ความนิยม คະแนนเสียง” มักปรากฏร่วมกับวลี *ช่วงมาก* ซึ่งเป็นบริบททางขวาเช่นกัน แต่สาเหตุที่เครื่องไม่เห็นความสัมพันธ์นี้น่าจะเนื่องมาจากตัวอย่างข้อมูลมีไม่เพียงพอ เพราะมีตัวอย่างเพียง 91 ตัวอย่าง ส่วนคำว่า *หัว* นั้นก็เป็นไปในทิศทางเดียวกัน คือ สิ่งที่จะช่วยบ่งชี้ความหมายของคำ ส่วนใหญ่มักเป็นบริบททางขวาเช่นกัน เช่น ความหมาย “ส่วนของพืชพันธุ์บางอย่างตอนที่อยู่ใต้ดิน” คำด้านขวาของ *หัว* จะเป็นชื่อพืช เช่น *มัน ผีอก* เป็นต้น ความหมาย “ส่วนแห่งสิ่งของบางอย่างที่อยู่ข้างหน้า หรือข้างต้น หรือแรกเริ่ม” คำปรากฏร่วมที่ปรากฏกับ *หัว* ได้แก่ *มุม ถนน เรือ เตี้ย* หรือความหมาย “ส่วนแห่งสิ่งของที่เป็นยอด”

คำที่ปรากฏรวม ได้แก่ ฝิ เช่า เป็นต้น จะเห็นได้ว่าบริบทที่เหมาะสมที่จะใช้บ่งชี้ความหมายของ หัว ยังคงเป็นบริบททางขวา

ในส่วนของคำกริยา สมมติฐานได้ตั้งไว้ว่าบริบทที่เหมาะสมในการช่วยแยกความหมายคือ บริบททางซ้ายและทางขวา ซึ่งจากการวิเคราะห์ความหมายตามแนวภาษาศาสตร์แล้วพบว่าขึ้นอยู่กับความหมายว่าเมื่อใดจะใช้บริบททางซ้ายหรือทางขวา เช่น คำว่า *บอก* ในความหมาย “พูดให้รู้, เล่าให้ฟัง” และ “บ่งให้รู้” สามารถแยกได้ด้วยบริบททางซ้าย กล่าวคือ ในความหมายแรก ผู้กระทำกริยานี้มักเป็นสิ่งมีชีวิต และความหมายที่สอง ผู้กระทำกริยามักเป็นสิ่งไม่มีชีวิต และคำว่า *ติด* ในความหมายว่า “อาการที่ข้องอยู่” มักปรากฏร่วมกับ *ยึด* ในบริบททางซ้าย แต่ก็มีบางตัวอย่างในความหมายนี้ที่ *ยึด* ปรากฏอยู่ทางขวา และความหมายว่า “ชอบอย่างขาดไม่ได้” พบว่าเป็นคำบริบททางซ้ายหรือขวา ก็ได้ หากเป็นบริบททางซ้ายจะปรากฏร่วมกับคำกริยา *เสพ* หากเป็นบริบททางขวาจะเป็นคำนาม เช่น *บุหรี่ยาเสพติด* เป็นต้น

ดังนั้นจึงสรุปได้ว่า บริบทที่เหมาะสมในการแยกความหมายของคำนามยังคงเป็นบริบททางขวา และบริบทที่เหมาะสมในการแยกความหมายของคำกริยาเป็นบริบททางซ้ายและขวา

6.3 ปัญหาที่พบ

ปัญหาที่พบในงานวิจัยนี้ จะเป็นปัญหาด้านตัวอย่างข้อมูลที่มีการเว้นวรรคตอนผิด เช่น

แต่เป็นการเชื่อมโยงด้วย||เสียง| เท่านั้น|

|เหรียญ| บน|หัว|ก็ใช้|กิป|ติด|ผม|ตัว|เรือน|ทอง|ฝัง|เพชร|

จากตัวอย่างแรก จะเห็นว่าทั้งก่อนหน้าและด้านหลังคำว่า *เสียง* ไม่จำเป็นต้องเว้นช่องว่างหรือในตัวอย่างที่ 2 *เหรียญ* ควรจะอยู่ติดกับวลีที่ตามมา ซึ่งจากการเว้นวรรคตอนผิดนี้ เมื่อระบบใช้ช่องว่างในการแบ่งกลุ่มความหมาย ก็ย่อมต้องมีผลต่อค่าความถูกต้องตามไปด้วยเนื่องจากจัดกลุ่มผิด

6.4 ข้อเสนอแนะ

เนื่องจากในงานวิจัยนี้ ใช้เพียงรูปคำในการแยกความหมายเป็นหลัก ดังนั้นแม้ระบบจะสามารถแยกความหมายของคำได้บ้าง แต่ก็ได้เฉพาะความหมายที่มีรูปแบบของคำปรากฏรวมที่ชัดเจนแตกต่างจากความหมายอื่น และมีความถี่สูงซึ่งในความเป็นจริงแล้วในภาษา คำบางคำสามารถเป็นได้หลายหน้าที่ ดังนั้นหากระบบไม่ได้มีข้อมูลทางภาษาในการช่วยแยกความหมาย ก็จะไม่สามารถแยกความแตกต่างของคำที่มีรูปเดียวกันแต่ต่างหน้าที่ได้ เช่น กรณีของคำว่า *ต่อ* ที่ระบบใช้ช่วยแยกนัยความหมายหนึ่งของหัว ซึ่ง *ต่อ* ที่ใช้แยกนี้ต้องเป็น *ต่อ* ที่ทำหน้าที่เป็นคำบุพบท หาก *ต่อ* ทำหน้าที่

เป็นคำกริยาก็จะใช้ไม่ได้ เช่น จุดต่อหัวกับลำตัว เป็นต้น ดังนั้น หากระบบได้ใช้ข้อมูลที่มีการกำกับหน้าที่ของหมวดคำด้วย ก็น่าจะทำให้ระบบได้เรียนรู้การแยกความหมายของคำได้ดีมากยิ่งขึ้น

นอกจากนี้ในงานวิจัยนี้ใช้จำนวนตัวอย่างในการเรียนรู้ของคำแต่ละคำไม่มาก ดังนั้นในบางความหมาย ระบบจึงอาจจะยังไม่เห็นรูปแบบของคำปรากฏรวมที่ชัดเจน เพราะจำนวนข้อมูลที่ให้เรียนรู้มีน้อยเกินไป เช่น ความหมายว่า “อยู่ในที่คุมขัง” มีเพียง 59 ตัวอย่างเท่านั้น ซึ่งในความเป็นจริงในความหมายนี้ คำที่มักปรากฏรวมกับว่า *ติด* จะเป็นคำว่า *คุก* หรือ *ตะราง* แต่เนื่องจากปริมาณข้อมูลที่ให้ระบบเรียนรู้มีน้อยเกินไป ทำให้ระบบไม่เห็นความแตกต่างของความหมายนี้จากความหมายอื่น ดังนั้นเพื่อเป็นการเพิ่มประสิทธิภาพของระบบ จึงควรเพิ่มปริมาณของข้อมูลที่ให้ระบบเรียนรู้ด้วย

ในงานวิจัยนี้แม้ว่าจะทำการเปรียบเทียบคำที่มีนัยความหมายเท่ากัน คือ *เสียง* และ *บอก* มี 3 ความหมาย และ *หัว* และ *ติด* มีความหมายเท่ากัน แต่ก็มีปัจจัยอื่นอีกที่ควรควบคุมด้วยคือการกระจายของข้อมูลในแต่ละความหมาย ที่ควรจะมีการกระจายที่ใกล้เคียงกัน ดังจะเห็นได้จากในงานวิจัยนี้ที่คำหนึ่งมีข้อมูลหนักไปที่บางความหมาย แต่อีกคำมีการกระจายกันของข้อมูล ก็อาจทำให้ค่าความถูกต้องที่ออกมาคลาดเคลื่อนได้

รายการอ้างอิง

ภาษาไทย

ราชบัณฑิตยสถาน. (Ed.) 2556. พจนานุกรม ฉบับราชบัณฑิตยสถาน พ.ศ. 2554 (2 ed.).

กรุงเทพฯ: ราชบัณฑิตยสถาน.

วิโรจน์ อรุณมานะกุล. 2545. Thai Word Segmentation. Retrieved from

<http://pioneer.chula.ac.th/~awirote/resources/thai-word-segmentation.html>

สุริยา รัตนกุล, ศาสตราจารย์ ดร. คุณหญิง. 2544. อรรถศาสตร์เบื้องต้น. สถาบันวิจัยภาษาและวัฒนธรรมเพื่อพัฒนาชนบท มหาวิทยาลัยมหิดล.

ภาษาอังกฤษ

Agirre, E. and Edmonds, P. 2007. Introduction. In Agirre, E. and Edmonds, P. (eds.).
Word Sense Disambiguation Algorithms and Application.

Clarke, D. 2007. Context-theoretic Semantics for Natural Language an Algebraic Framework. Doctor of Philosophy's thesis. University of Sussex.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., and Harshman, R. 1990.
Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science 41(6), pp. 391-407.

Dumais, S.T. 1991. Improving the retrieval of information from external sources.
Behavior Research Methods, Instruments, & Computers, 23 (2).

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence.
Computational Linguistics 19(1).

Edmonds, P. 2006. Disambiguation, Lexical. In Encyclopedia of language and linguistics (2nd ed). Elsevier.

Evans, V. 2004. The Structure of Time: Language, Meaning and Temporal Cognition.
Amsterdam: John Benjamins. Cited in Evans, V. and Green, M. 2006. Cognitive Linguistics An Introduction. Edinburgh: Edinburgh University Press.

- Evans, V. and Tyler, A. 2003. Towards a Theory of Principled Polysemy: The Case of In. ICLC 2003.
- Evans, V. and Green, M. 2006. Cognitive Linguistics An Introduction. Edinburgh: Edinburgh University Press.
- Falkum, I.L. 2011. The Semantics and Pragmatics of Polysemy: A Relevance-Theoretic Account. Doctor of Philosophy's thesis.University College London.
- Falkum, I.L., and Vicente, A. 2015. Polysemy: Current Perspectives and Approaches. Lingua, Volume 157.
- Fillmore, J. and Atkins, B.T.S. 2006. Describing Polysemy. In Ravin, Y. and Leacock, C. (eds.) Polysemy: Theoretical and Computational Approaches. New York: Oxford University Press.
- Firth, J.R. 1957. Papers in Linguistics (1934-1951).London: Oxford University Press.
- Gella, S., Cook, P., and Han, B. 2013.Unsupervised Word Usage Similarity in Social Media Texts. In Second Joint Conference on Lexical and Computational Semantics, Vol.2: Seventh International Workshop on Semantic Evaluation (SemEval 2013). Atlanta, Georgia.
- Harris, Z. 1968. Mathematical Structures of Language. New York.
- Ide, N. and Véronis, J. 1998. Computational Linguistics 24(1).Special issue on Word Sense Disambiguation.Cited in Edmonds, P. 2006.Disambiguation, Lexical. In Encyclopedia of language and linguistics (2nd ed). Elsevier.
- Jurafsky, D. and Martin, J.H. 2009. Speech and Language Processing. Pearson.
- Kanokrattananukul, W. (2001). *Word Sense Disambiguation in Thai Using Decision List Collocation*.(Master of Arts Degree Thesis, Linguistics) Chulalongkorn University, Bangkok.
- Kontostathis, A., and Pottenger, W.M. 2002. Detecting patterns in the LSI term-term matrix. In Proceedings of the ICDM 2002 Workshop on Foundations of Data Mining and Discovery.

- Landauer, T.K., Foltz, P.W., and Laham, D. 1998. Introduction to Latent Semantic Analysis. Discourse Processes, 25.
- Landauer, T.K., Laham, D., and Foltz, P. 1998a. Learning Human-like Knowledge by Singular Value Decomposition: A Progress Report. In Report, M.I., Jordan, M.J., Kearns & S.A. Sollar (eds.). Advances in Neural Information Processing Systems 10. Cambridge: MIT Press.
- Levin, E., Sharifi, M., and Ball, J. 2006. Evaluation of Utility of LSA for Word Sense Discrimination. In Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL. New York.
- Lin, D. 1998. Automatic Retrieval and Clustering of Similar Words. In Proceedings of the 17th Int'l Conference on Computational Linguistics. Quebec, Canada. Cited in Nasiruddin, M. 2013. A State of the Art of Word Sense Induction: A Way Towards Word Sense Disambiguation for Under-Resourced Languages.
- Lloyd, S.P. 1982. Least squares quantization in PCM. In *IEEE Transactions on Information Theory*, 28 (2): 129–137.
- Lyons, J. 1996. Linguistic Semantics: An Introduction. Cambridge: Cambridge University Press.
- Màrquez, L., Escudero, G., Martínez, D., and Rigau, G. 2007. Supervised Corpus-Based Methods for WSD. In Agirre, E. and Edmonds, P. (eds.). Word Sense Disambiguation Algorithms and Application.
- Mihalcea, R., 2007. Knowledge-Based Methods for WSD. In Agirre, E. and Edmonds, P. (eds.). Word Sense Disambiguation Algorithms and Application.
- Modhiran, T., Kruatrachue, B., and Supnithi, T. 2004. Comparison Thai Word-Sense Disambiguation Method. In International Conference on Control, Automation and Systems.
- Murphy, M.L. 2010. Lexical Meaning. Cambridge: Cambridge University Press.
- Nasiruddin, M. 2013. A State of the Art of Word Sense Induction: A Way Towards Word Sense Disambiguation for Under-Resourced Languages.

- Navigli, R. 2009. Word Sense Disambiguation: A Survey. In ACM Computing Surveys, Vol.42, No.2.
- _____. 2012. A Quick Tour of Word Sense Disambiguation, Induction and Related Approaches. In Bieliková, M., Friedrich, G., Gottlob, G., Katzenbeisser, S., and Turán, G. (eds.). SOFSEM 2012. LNCS, vol.7147, Heidelberg.
- Palmer, F.R. 1981. Semantics (2nd ed). Cambridge: Cambridge University Press.
- Pedersen, T. 2007. Unsupervised Corpus-Based Methods for WSD. In Agirre, E. and Edmonds, P. (eds.). Word Sense Disambiguation Algorithms and Application.
- Pino, J. and Eskenazi, M. 2009. An Application of Latent Semantic Analysis to Word Sense Discrimination for Words with Related and Unrelated Meanings. In Proceedings of the NSSCL HLT Workshop on Innovation Use of NLP for Building Educational Applications. Boulder, Colorado.
- Pongpinigpinyo, S. and Rivepiboon, W. 2005. Distributional Semantic Approach to Thai Word Sense Disambiguation. In International Journal of Computational Intelligence Vol.2 No.3 2005.
- Rada, R., Mili, H., Bicknell, E., and Blettner, M. 1989. Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man, and Cybernetics, 19(1). Cited in Mihalcea, R., 2007. Knowledge-Based Methods for WSD. In Agirre, E. and Edmonds, P. (eds.). Word Sense Disambiguation Algorithms and Application.
- Ravin, Y. and Leacock, C. 2006. Polysemy: An Overview. In Ravin, Y. and Leacock, C. (eds.) Polysemy: Theoretical and Computational Approaches. New York: Oxford University Press.
- Riemer, N. 2005. The Semantics of Polysemy: Reading Meaning in English and Warlpiri. Berlin: Mouton de Gruyter. อ้างถึงใน จิรัชย์ หิรัญรัต. 2550. การศึกษาความหมายของคำว่า 'เอา' ในภาษาไทย. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต, ภาควิชาภาษาศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- Rosch, E. 1977. Human categorization. In Warren, N. (ed.). Advances in Cross-Cultural Psychology, vol. 7. London: Academic Press. Cited in Ravin, Y. and Leacock, C.

- (eds.) Polysemy: Theoretical and Computational Approaches. New York: Oxford University Press.
- Saeed, J.I. 2004. Semantics (2nd ed). Second edition. Oxford: Blackwell.
- Sandra, D. 1998. What linguists can and can't tell you about the human mind: a reply to Croft. In Cognitive Linguistics, 9, 4. Cited in Evans, V. and Green, M. 2006. Cognitive Linguistics An Introduction. Edinburgh: Edinburgh University Press.
- Schütze, H. 1992. Dimensions of meaning. In Supercomputing'92: Proceedings of the 1992 ACM/IEEE Conference on Supercomputing. IEEE Computer Society Press, Los Alamitos, CA. Cited in Navigli, R. 2009. Word Sense Disambiguation: A Survey. In ACM Computing Surveys, Vol.42, No.2.
- _____. 1998. Automatic Word Sense Discrimination. In Computational Linguistics 24(1).MIT Press.
- Tan, P., Steinbach, M., and Kumar, V. 2005. Introduction to Data Mining. Massachusetts: Addison-Wesley Longman Publishing.
- Tyler, A. and Evans, V. 2003. The Semantics of English Prepositions: Spatial Scenes. Embodied Meaning and Cognition. Cambridge: Cambridge University Press. Cited in Evans, V. and Green, M. 2006. Cognitive Linguistics An Introduction. Edinburgh: Edinburgh University Press.

ภาคผนวก

รายละเอียดการจัดกลุ่มความหมายของระบบของ *เสียง บอก หัว ดิด*

ส่วนที่เป็นคอลัมน์ คือ กลุ่มตัวอย่างข้อมูลที่เครื่องแยกออกเป็นกลุ่ม ๆ ส่วนที่เป็นแถวคือความหมายที่วิเคราะห์ได้ ส่วนที่แรเงาคือคำตอบที่โปรแกรมเลือกให้เป็นคำตอบของความหมายนั้น ๆ

เสียง

ตาราง 1 รายละเอียดการจัดกลุ่มความหมายของ *เสียง* ด้วยคำบริบททางซ้าย 2 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย / ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. สิ่งที่ได้รับรู้ได้ด้วยหู	1698	89	87	1874
2. ความเห็น	51	1	1	53
3. คะแนนเสียง	89		2	91
รวม				2018

ตาราง 2 รายละเอียดการจัดกลุ่มความหมายของ *เสียง* ด้วยคำบริบททางซ้าย 3 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย / ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. สิ่งที่ได้รับรู้ได้ด้วยหู	20	1709	145	1874
2. ความเห็น	1	51	1	53
3. คะแนนเสียง	1	86	4	91
รวม				2018

ตาราง 3 รายละเอียดการจัดกลุ่มความหมายของ *เสียง* ด้วยคำบริบททางซ้าย 5 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย / ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. สิ่งที่ได้รับรู้ได้ด้วยหู	91	1736	47	1874
2. ความเห็น		51	2	53
3. คะแนนเสียง		89	2	91

รวม	2018
-----	------

ตาราง 4 รายละเอียดการจัดกลุ่มความหมายของ *เสียง* ด้วยคำบริบททางขวา 2 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. สิ่งที่ได้รับรู้ได้ด้วยหู	150	1673	51	1874
2. ความเห็น		48	5	53
3. คะแนนเสียง		74	17	91
รวม				2018

ตาราง 5 รายละเอียดการจัดกลุ่มความหมายของ *เสียง* ด้วยคำบริบททางขวา 3 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. สิ่งที่ได้รับรู้ได้ด้วยหู	202	186	1486	1874
2. ความเห็น		8	45	53
3. คะแนนเสียง		8	83	91
รวม				2018

ตาราง 6 รายละเอียดการจัดกลุ่มความหมายของ *เสียง* ด้วยคำบริบททางขวา 5 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. สิ่งที่ได้รับรู้ได้ด้วยหู	202	186	1486	1874
2. ความเห็น		8	45	53
3. คะแนนเสียง		8	83	91
รวม				2018

ตาราง 7 รายละเอียดการจัดกลุ่มความหมายของ *เสียง* ด้วยคำบริบททางซ้าย-ขวา 2 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. สิ่งที่ได้รับรู้ได้ด้วยหู	1874	285	1089	1,874
2. ความเห็น	53	22	23	53
3. คະแนนเสียง	91	23	62	91
รวม				2018

ตาราง 8 รายละเอียดการจัดกลุ่มความหมายของ *เสียง* ด้วยคำบริบททางซ้าย-ขวา 3 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. สิ่งที่ได้รับรู้ได้ด้วยหู	634	1081	159	1874
2. ความเห็น	19	29	5	53
3. คະแนนเสียง	18	47	26	91
รวม				2018

ตาราง 9 รายละเอียดการจัดกลุ่มความหมายของ *เสียง* ด้วยคำบริบททางซ้าย-ขวา 5 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. สิ่งที่ได้รับรู้ได้ด้วยหู	783	939	152	1874
2. ความเห็น	22	30	1	53
3. คະแนนเสียง	38	48	5	91
รวม				2018

บอก

ตาราง 10 รายละเอียดการจัดกลุ่มความหมายของ บอก ด้วยคำบริบททางซ้าย 2 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	1421	80	138	1639
2. บ่งให้รู้	133	7	28	168
3. แนะนำ, สอน	184	4	23	211
รวม				2018

ตาราง 11 รายละเอียดการจัดกลุ่มความหมายของ บอก ด้วยคำบริบททางซ้าย 3 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	104	301	1234	1639
2. บ่งให้รู้	7	35	126	168
3. แนะนำ, สอน	11	47	153	211
รวม				2018

ตาราง 12 รายละเอียดการจัดกลุ่มความหมายของ บอก ด้วยคำบริบททางซ้าย 5 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	738	532	369	1639
2. บ่งให้รู้	86	56	26	168
3. แนะนำ, สอน	92	64	55	211
รวม				2018

ตาราง 13 รายละเอียดการจัดกลุ่มความหมายของ *บอก* ด้วยคำบริบททางขวา 2 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	775	749	115	1639
2. บ่งให้รู้	47	115	6	168
3. แนะนำ, สอน	60	141	10	211
รวม				2018

ตาราง 14 รายละเอียดการจัดกลุ่มความหมายของ *บอก* ด้วยคำบริบททางขวา 3 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	822	712	105	1639
2. บ่งให้รู้	66	93	9	168
3. แนะนำ, สอน	67	142	2	211
รวม				2018

ตาราง 15 รายละเอียดการจัดกลุ่มความหมายของ *บอก* ด้วยคำบริบททางขวา 5 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	1031	365	243	1639
2. บ่งให้รู้	84	58	26	168
3. แนะนำ, สอน	74	25	112	211
รวม				2018

ตาราง 16 รายละเอียดการจัดกลุ่มความหมายของ *บอก* ด้วยคำบริบททางซ้าย-ขวา 2 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	825	644	170	1639
2. บ่งให้รู้	49	84	35	168
3. แนะนำ, สอน	64	55	92	211
รวม				2018

ตาราง 17 รายละเอียดการจัดกลุ่มความหมายของ *บอก* ด้วยคำบริบททางซ้าย-ขวา 3 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	942	521	176	1639
2. บ่งให้รู้	76	79	13	168
3. แนะนำ, สอน	71	72	68	211
รวม				2018

ตาราง 18 รายละเอียดการจัดกลุ่มความหมายของ *บอก* ด้วยคำบริบททางซ้าย-ขวา 5 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	รวม
1. พูดให้รู้, เล่าให้ฟัง	374	578	687	1639
2. บ่งให้รู้	30	83	55	168
3. แนะนำ, สอน	31	132	48	211
รวม				2018

หัว

ตาราง 19 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางซ้าย 2 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. ส่วนบนสุดของ ร่างกายของคนหรือสัตว์	884	27		14	6	16	14	30	22	1013
2. ส่วนของพืชพันธุ์ บางอย่างตอนที่อยู่ใต้ ดิน	64					2	1	2	1	70
3. ส่วนแห่งสิ่งของ บางอย่างที่อยู่ข้างหน้า หรือข้างต้น	230			8		2	4	9	1	254
4. ส่วนแห่งสิ่งของที่เป็น ยอด	141			6		1		9	1	158
5. ผู้ที่มีความคิดหนักไป ทางใดทางหนึ่ง	54							2		56
6. ปัญญา, ความคิด	165			18		1	1	4	2	191
7. ผม	86		1						2	89
8. นามนัยแทนสิ่งที่ กล่าวถึงทั้งหมด	68		38							106
9. ส่วนของอุปกรณ์	61					3		5		69
รวม										2006

ตาราง 20 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางซ้าย 3 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. ส่วนบนสุดของ ร่างกายของคนหรือสัตว์	27	16	42	18	16		93	50	751	1013
2. ส่วนของพืชพันธุ์ บางอย่างตอนที่อยู่ใต้ ดิน		2	3		1		3	3	58	70

3. ส่วนแห่งสิ่งของ บางอย่างที่อยู่ข้างหน้า หรือข้างต้น		2	2	12	6		15	33	184	254
4. ส่วนแห่งสิ่งของที่ เป็นยอด		2	2	8	3		12	14	117	158
5. ผู้ที่มีความคิดหนักไป ทางใดทางหนึ่ง			1					4	51	56
6. ปัญญา, ความคิด		5	1	21	1		30	10	123	191
7. ผม		3	4		1	1	5	6	69	89
8. นามนัยแทนสิ่งที่ กล่าวถึงทั้งหมด		2			1	39		3	61	106
9. ส่วนของอุปกรณ์							3	8	58	69
รวม										2006

ตาราง 21 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางซ้าย 5 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. ส่วนบนสุดของ ร่างกายของคนหรือสัตว์	27	607	35	7	115			130	92	1013
2. ส่วนของพืชพันธุ์ บางอย่างตอนที่อยู่ได้ ดิน		46	7		7			6	4	70
3. ส่วนแห่งสิ่งของ บางอย่างที่อยู่ข้างหน้า หรือข้างต้น		144	13	2	16			32	47	254
4. ส่วนแห่งสิ่งของที่ เป็นยอด		90	11	5	10	1		18	23	158
5. ผู้ที่มีความคิดหนักไป ทางใดทางหนึ่ง		30	14		3			3	6	56
6. ปัญญา, ความคิด		102	14		24	1		28	22	191
7. ผม		50	4	1	9			15	10	89
8. นามนัยแทนสิ่งที่ กล่าวถึงทั้งหมด		30	25		3		37	6	5	106
9. ส่วนของอุปกรณ์		39	4		7			7	12	69

รวม	2006
-----	------

ตาราง 22 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางขวา 2 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. ส่วนบนสุดของ ร่างกายของคนหรือสัตว์	830		41	44	29	44	24		1	1013
2. ส่วนของพืชพันธุ์ บางอย่างตอนที่อยู่ใต้ ดิน	63		2	2	2	1				70
3. ส่วนแห่งสิ่งของ บางอย่างที่อยู่ข้างหน้า หรือข้างต้น	194		6	19	5	9	4	1	16	254
4. ส่วนแห่งสิ่งของที่ เป็นยอด	113		5	7	3	2		28		158
5. ผู้ที่มีความคิดหนักไป ทางใดทางหนึ่ง	50		1	3	2					56
6. ปัญญา, ความคิด	159		23	4	3	2				191
7. ผม	85		3	1						89
8. นามนัยแทนสิ่งที่ กล่าวถึงทั้งหมด	87		12	3	4					106
9. ส่วนของอุปกรณ์	61	2	1	2		1			2	69
รวม										2006

ตาราง 23 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางขวา 3 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. ส่วนบนสุดของ ร่างกายของคนหรือสัตว์	37	22	49	715	7	55	25	39	64	1013
2. ส่วนของพืชพันธุ์ บางอย่างตอนที่อยู่ใต้ ดิน	4	3	3	51		5		3	1	70

3. ส่วนแห่งสิ่งของ บางอย่างที่อยู่ข้างหน้า หรือข้างต้น	13	11	8	176	1	22	4	9	10	254
4. ส่วนแห่งสิ่งของที่ เป็นยอด	5	3	6	126		11		5	2	158
5. ผู้ที่มีความคิดหนักไป ทางใดทางหนึ่ง	5	2	1	44		3		1		56
6. ปัญญา, ความคิด	5	12	23	138		5		3	5	191
7. ผม	6	6	4	71		2				89
8. นามนัยแทนสิ่งที่ กล่าวถึงทั้งหมด	12	2	15	67		3		6	1	106
9. ส่วนของอุปกรณ์	2	2	1	55		4		3	2	69
รวม										2006

ตาราง 24 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางขวา 5 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. ส่วนบนสุดของ ร่างกายของคนหรือสัตว์	61	100	59	72	126	477		93	25	1013
2. ส่วนของพืชพันธุ์ บางอย่างตอนที่อยู่ได้ ดิน	7	2	3	6	10	39		3		70
3. ส่วนแห่งสิ่งของ บางอย่างที่อยู่ข้างหน้า หรือข้างต้น	22	7	9	30	33	135	1	13	4	254
4. ส่วนแห่งสิ่งของที่ เป็นยอด	27	3	8	15	17	80	2	6		158
5. ผู้ที่มีความคิดหนักไป ทางใดทางหนึ่ง	10	3	2	6	10	25				56
6. ปัญญา, ความคิด	7	12	26	6	32	94		14		191
7. ผม	9	11	4	2	11	48		4		89
8. นามนัยแทนสิ่งที่ กล่าวถึงทั้งหมด	14	9	14	5	19	42		3		106
9. ส่วนของอุปกรณ์	2	1	1	2	4	20	37	2		69

รวม	2006
-----	------

ตาราง 25 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางซ้าย-ขวา 2 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. ส่วนบนสุดของ ร่างกายของคนหรือสัตว์	111	27		1	45	713	29	87		1013
2. ส่วนของพืชพันธุ์ บางอย่างตอนที่อยู่ใต้ ดิน	4				1	61		4		70
3. ส่วนแห่งสิ่งของ บางอย่างที่อยู่ข้างหน้า หรือข้างต้น	75				17	145	4	13		254
4. ส่วนแห่งสิ่งของที่ เป็นยอด	23				6	123		6		158
5. ผู้ที่มีความคิดหนักไป ทางใดทางหนึ่ง	6				1	48		1		56
6. ปัญญา, ความคิด	8				10	146		27		191
7. ผม	7		1		7	71		3		89
8. นามนัยแทนสิ่งที่ กล่าวถึงทั้งหมด	2		38	16	2	20			28	106
9. ส่วนของอุปกรณ์	10					57		2		69
รวม										2006

ตาราง 26 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางซ้าย-ขวา 3 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. ส่วนบนสุดของ ร่างกายของคนหรือสัตว์	77	12	27	41	34	50	281	484	7	1013
2. ส่วนของพืชพันธุ์ บางอย่างตอนที่อยู่ใต้ ดิน	6				1	3	12	48		70

3. ส่วนแห่งสิ่งของ บางอย่างที่อยู่ข้างหน้า หรือข้างต้น	39			7	7	6	62	133		254
4. ส่วนแห่งสิ่งของที่ เป็นยอด	17	1		2	6	7	27	98		158
5. ผู้ที่มีความคิดหนักไป ทางใดทางหนึ่ง	9				1		16	30		56
6. ปัญญา, ความคิด	14	7				4	101	65		191
7. ผม	11			1	2	3	21	51		89
8. นามนัยแทนสิ่งที่ กล่าวถึงทั้งหมด	75			1			7	23		106
9. ส่วนของอุปกรณ์	10				1	3	16	39		69
รวม										2006

ตาราง 27 รายละเอียดการจัดกลุ่มความหมายของ หัว ด้วยคำบริบททางซ้าย-ขวา 5 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. ส่วนบนสุดของ ร่างกายของคนหรือสัตว์	94	102	175	338	178	27	94		5	1013
2. ส่วนของพืชพันธุ์ บางอย่างตอนที่อยู่ได้ ดิน	12	3	18	18	9		10			70
3. ส่วนแห่งสิ่งของ บางอย่างที่อยู่ข้างหน้า หรือข้างต้น	38	32	48	88	26		22			254
4. ส่วนแห่งสิ่งของที่ เป็นยอด	23	4	43	44	30		14			158
5. ผู้ที่มีความคิดหนักไป ทางใดทางหนึ่ง	20	2	7	12	7		8			56
6. ปัญญา, ความคิด	24	36	16	36	49		30			191
7. ผม	8	11	19	31	14		5	1		89
8. นามนัยแทนสิ่งที่ กล่าวถึงทั้งหมด	30	1	14	10	6		6	39		106
9. ส่วนของอุปกรณ์	14	5	20	15	7		8			69

รวม	2006
-----	------

ติด

ตาราง 28 รายละเอียดการจัดกลุ่มความหมายของ ติด ด้วยคำบริบททางซ้าย 2 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. อาการที่ข้ออยู่	10	5	1	123		2	51		21	213
2. ทำให้เกิดขัดอยู่หรือ ข้ออยู่จนไม่อาจให้ เคลื่อนไปได้	19	3	23	198		8	19		33	303
3. ขอบอย่างขาดไม่ได้	4	2		127	36	1	7		23	200
4. ผนึก	7			182		3	10		25	227
5. ต่อเนื่องอย่างใกล้ชิด	7		1	261		80	3	38	45	435
6. ได้รับเชื้อโรค	9	6	1	251	1	2	20		18	308
7. คงอยู่หรือทำให้คง อยู่	10	2		147		3	7	2	30	201
8. อยู่ในที่คุมขัง	1	1		47		1	2		7	59
9. ผ่านเกณฑ์	4		1	40		2	8		5	60
รวม										2006

ตาราง 29 รายละเอียดการจัดกลุ่มความหมายของ ติด ด้วยคำบริบททางซ้าย 3 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. อาการที่ข้ออยู่	53	32		4	1	7		112	4	213
2. ทำให้เกิดขัดอยู่หรือ ข้ออยู่จนไม่อาจให้ เคลื่อนไปได้	49	39		5	1	4		195	10	303
3. ขอบอย่างขาดไม่ได้	12	24	36	5		3	2	113	5	200
4. ผนึก	11	29		1	2	7		163	14	227
5. ต่อเนื่องอย่างใกล้ชิด	5	49		3	49	9		304	16	435

6. ได้รับเชื้อโรค	13	24	1	53		7		180	30	308
7. คงอยู่หรือทำให้คงอยู่	6	32		1	3	11		122	26	201
8. อยู่ในที่คุมขัง	3	8		1		3		41	3	59
9. ผ่านเกณฑ์	8	9		2		4		34	3	60
รวม										2,006

ตาราง 30 รายละเอียดการจัดกลุ่มความหมายของ คิต ด้วยคำบริบททางซ้าย 5 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. อาการที่ช่องอยู่	18	13	17	5	85	19	12		44	213
2. ทำให้เกิดขัดอยู่หรือ ช่องอยู่จนไม่อาจให้ เคลื่อนไปได้	36	34	10	13	131	17	3		59	303
3. ขอบอย่างขาดไม่ได้	24	14	17	7	69	9	6	38	16	200
4. ผนึก	22	27	16	19	113	14	5		11	227
5. ต่อเนื่องอย่างใกล้ชิด	41	44	50	17	239	23	12		9	435
6. ได้รับเชื้อโรค	41	22	67	36	102	9	11	1	19	308
7. คงอยู่หรือทำให้คง อยู่	19	29	8	29	86	8	8		14	201
8. อยู่ในที่คุมขัง	8	10	3	3	23	1	3		8	59
9. ผ่านเกณฑ์	6	8	3	4	26	2	1		10	60
รวม										2006

ตาราง 31 รายละเอียดการจัดกลุ่มความหมายของ คิต ด้วยคำบริบททางขวา 2 ตำแหน่ง

ผลการจัดกลุ่มของ ความหมาย ระบบ	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. อาการที่ช่องอยู่	140	32	1	18		4			18	213
2. ทำให้เกิดขัดอยู่หรือ ช่องอยู่จนไม่อาจให้ เคลื่อนไปได้	216	2	1	62		2	2		18	303
3. ขอบอย่างขาดไม่ได้	177		17	1	1	2	2			200
4. ผนึก	205	3		17		2				227

5. ต่อเนื่องอย่างใกล้ชิด	263	10		4		87	46	25		435
6. ได้รับเชื้อโรค	34		1	1	270	2				308
7. คงอยู่หรือทำให้คงอยู่	134	9	1	50		1			6	201
8. อยู่ในที่คุมขัง	52			6					1	59
9. ผ่านเกณฑ์	56			1					3	60
รวม										2006

ตาราง 32 รายละเอียดการจัดกลุ่มความหมายของ คิต ด้วยคำบริบททางขวา 3 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. อาการที่ซ้องอยู่		71	5	1		49		84	3	213
2. ทำให้เกิดขัดอยู่หรือ ซ้องอยู่จนไม่อาจให้ เคลื่อนไปได้	11	84	6	1		7		189	5	303
3. ซอบอย่างขาดไม่ได้	4	2	7		1		49	114	23	200
4. ผนึก		21	5	34		13		153	1	227
5. ต่อเนื่องอย่างใกล้ชิด		15	150	1		140		127	2	435
6. ได้รับเชื้อโรค	1		4		273		1	29		308
7. คงอยู่หรือทำให้คง อยู่	7	65	3	9		28		88	1	201
8. อยู่ในที่คุมขัง	2	8	1					46	2	59
9. ผ่านเกณฑ์		5	1					54		60
รวม										2006

ตาราง 33 รายละเอียดการจัดกลุ่มความหมายของ คิต ด้วยคำบริบททางขวา 5 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. อาการที่ซ้องอยู่	1	71	63	51		3		4	20	213
2. ทำให้เกิดขัดอยู่หรือ ซ้องอยู่จนไม่อาจให้ เคลื่อนไปได้	12	87	141	9		13		4	37	303
3. ซอบอย่างขาดไม่ได้	7	4	110	1	1	5	51	7	14	200

4. ผนัง	1	21	139	14		5		4	43	227
5. ต่อเนื่องอย่างใกล้ชิด	3	18	107	144		39		103	21	435
6. ได้รับเชื้อโรค	2		23		273	2	1	3	4	308
7. คงอยู่หรือทำให้คงอยู่	13	66	75	30		1		2	14	201
8. อยู่ในที่คุมขัง	2	8	45			1		1	2	59
9. ผ่านเกณฑ์		5	41			2			12	60
รวม										2006

ตาราง 34 รายละเอียดการจัดกลุ่มความหมายของ ดิต ด้วยคำบริบททางซ้าย-ขวา 2 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. อาการที่ข้ออยู่	25	65	49		14	3		57		213
2. ทำให้เกิดขัดอยู่หรือ ข้ออยู่จนไม่อาจให้ เคลื่อนไปได้	119	89	42		1	25		27		303
3. ขอบอย่างขาดไม่ได้	92	1	14	1	32	2	36	22		200
4. ผนัง	125	20	12		6	20		44		227
5. ต่อเนื่องอย่างใกล้ชิด	152	83	3		2	62		133		435
6. ได้รับเชื้อโรค	22		3	120	158		1	4		308
7. คงอยู่หรือทำให้คง อยู่	73	65	5	1	3	18		36		201
8. อยู่ในที่คุมขัง	38	7	2		7	2		3		59
9. ผ่านเกณฑ์	31	4	7		1			8	9	60
รวม										2006

ตาราง 35 รายละเอียดการจัดกลุ่มความหมายของ ดิต ด้วยคำบริบททางซ้าย-ขวา 3 ตำแหน่ง

ผลการจัดกลุ่มของ ระบบ ความหมาย	กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่ม 4	กลุ่ม 5	กลุ่ม 6	กลุ่ม 7	กลุ่ม 8	กลุ่ม 9	รวม
1. อาการที่ข้ออยู่	31	71	11	1		56		12	31	213
2. ทำให้เกิดขัดอยู่หรือ ข้ออยู่จนไม่อาจให้ เคลื่อนไปได้	105	92	5	6		9		41	45	303

ประวัติผู้เขียนวิทยานิพนธ์

นางสาว นัชชา ถิระสาโรช เกิดที่จังหวัดเชียงใหม่ สำเร็จการศึกษาระดับปริญญาตรี ศิลปศาสตรบัณฑิต สาขาภาษาอังกฤษ มหาวิทยาลัยเชียงใหม่ ในปีการศึกษา 2544 ระดับปริญญาโท อักษรศาสตรมหาบัณฑิต ภาควิชาภาษาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2553 และเข้าศึกษาต่อในหลักสูตรปริญญาอักษรศาสตรดุษฎีบัณฑิต สาขาวิชาภาษาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2554